Vol.2018-DPS-175 No.19
Vol.2018-MBL-87 No.19
Vol.2018-ITS-73 No.19
2018/5/24

モバイルネットワーク特徴量を用いた Contextual Bandit Algorithm

出水宰[†1]　Rubén Manzano[†2]　Sergio Gómez[†2]　深澤佑介[†1]

概要：モバイルサービス上の広告の送客を増やすための手法として，ユーザの特徴量を考慮してコンテンツを選択する Contextual bandit algorithm がある．しかしながら，モバイルサービス側でユーザに対する十分な特徴量を常に取得できているとは限らない．本論文では，どのようなモバイルサービスでも汎用的に Contextual bandit algorithm を適用できるようにすることを目的として，モバイルネットワークに基づく特徴量のみを利用し，次元圧縮とクラスタリングによる動的なアプローチを提案する．スペインとギリシャのキャンペーンサイトでのデータセットに提案手法を適用した検証において，スペインでは A/B テストから 9%，特徴量を考慮しない Bandit algorithm から 4%のコンバージョン率向上を確認した．

1. はじめに

Web 上での広告配信においては，コンバージョン効果の高いコンテンツを選択することは重要である．例えばキャンペーンサイトのランディングページ上に，複数ある広告クリエイティブの中からどれを選択するかにより，クリック数やコンバージョン数は変わってくる．しかし，意思決定者は，それらの広告クリエイティブのクリック率やコンバージョン率を事前に知ることはできないため，表示のテストを行いながら有効なものを選択する必要がある．

この行動選択の際に，意思決定者は大きく二つの戦略を取ることが可能である．一つは，探索（exploration）と呼ばれる戦略で，複数ある広告クリエイティブをランダムに表示させることにより，それぞれのクリック率の期待値を得ることができる．もう一方は，活用（exploitation）と呼ばれる戦略で，今判明している推定のクリック率の中で，最も値が高い広告クリエイティブを選択することにより，得られる報酬を増やすことができる．これらはトレードオフの関係にあり，探索（exploration）を重視しすぎるとランダム性が強くなり，真に効果の高い広告クリエイティブを選択する機会が減ってしまい，その結果，報酬を増やすことができなくなる．同様に，活用（exploitation）を重視しすぎると，広告クリエイティブの推定クリック率の学習が進んでいない中で，誤った選択になる可能性がある．

こうしたトレードオフの関係をバランスさせるアプローチとして，Bandit algorithm[1]がよく用いられる． Bandit algorithm では，一定期間における累計のクリック数やコンバージョン数といった報酬の最大化を目的として，exploitation と exploitation をバランスさせながら，広告クリエイティブを選択していく． この Bandit algorithm の拡張として，ユーザの特徴量に応じて有効な広告クリエイティブを出し分けることが考えられ， Contextual bandit algorithm[21]と呼ばれている．コンテキストを考慮しない Bandit algorithm では，全てのユーザに対して同じ期待報酬の分布を仮定している（以降，コンテキストを考慮しない Bandit algorithm を Context-free bandit algorithm と呼ぶ）．これに対して Contextual bandit algorithm では，それぞれの広告クリエイティブが特徴量に応じた期待報酬の分布を持つと仮定している．そのため，ユーザごとによりマッチしたコンテンツを提供することができるため, Context-fee bandit algorithm に比べて報酬をさらに増やすことができる．

Contextual bandit algorithms のサービス適用としては，ニュース記事の推薦や広告表示のパーソナライズなどに適用されている．ニュース記事の推薦の例では，Web ページに表示する記事の出し分けの際に，ユーザの特徴量を考慮したことにより，通常の Bandit algorithm に比べて 12.5%のクリック持ち上げ効果を達成した[10].

こうした Contextual bandit algorithms で用いられるユーザの特徴量に着目した場合，特徴量の種類は大きく以下のように分類することができる．

- Demographic and geographic features: 性別・年代や住居エリアといった属性情報など
- Behavioral features: サービスの利用ログなど
- Implicit features: 端末や通信状況など

過去の適用例では、広告クリエイティブのクリックと相関があると思われる Demographic and geographic features や Behavioral features をコンテキスト情報として利用するケースが多い．これらのログは獲得コストが高い分，レコメンドの精度向上のために大きく寄与すると思われる. 一方で，端末情報や通信状況といった Implicit features のコンテキスト情報は，広告クリエイティブのクリックとの相関は明確ではないものの，獲得コストは非常に容易である．

Contextual bandit algorithms のサービスへの適用を想定した際に，ユーザに対する十分な特徴量（Demographic and geographic features や Behavioral features）がサービス側で常に得られるとは限らない．そこで本論文では，この Implicit features の情報であるモバイルネットワーク特徴量に着目した Contextual bandit algorithm を提案する．本論文の貢献

†1 株式会社ＮＴＴドコモ
NTT DOCOMO, INC.
†2 DOCOMO Digital Limited

Vol.2018-DPS-175 No.19
Vol.2018-MBL-87 No.19
Vol.2018-ITS-73 No.19
2018/5/24

内容は，次の通りである．

- スパースなモバイルネットワーク特徴量を活用するため，特徴量の次元圧縮とクラスタリング処理をContextual bandit algorithm に導入した．
- ユーザのコンテキストを収集する期間を考慮し，Context-free bandit と Contextual bandit の併用アルゴリズムを提案した．
- モバイルサービスにおける実データでオフライン検証を行い，獲得コストの低いネットワーク特徴量のみで Context-free bandit からの性能向上を確認した．

本稿の構成は以下のようである．第 2 章では，Contextual bandit algorithm に関連する研究について紹介する．第 3 章では，対象のサービスや問題設定について定義する．第 4 章では，提案手法の次元圧縮とクラスタリングを活用した Contextual bandit algorisms について述べる． 第 5 章では，オフライン環境での提案手法の性能検証について説明する．

2. 関連研究

本章では， Bandit algorithm，Contextual bandit algorithm 及び，そのサービス適用事例について紹介する．

2.1 Bandit algorithm

Bandit algorithm では，得られる報酬の最大化を目的として，exploitation と exploitation をバランスさせながら行動を選択していく．このバランスを決めるポリシーの種類には，ε-greedy[3]，Softmax[4]，UCB1[2]，Exp3[5]や， Thompson sampling[6][7][8][9]などがある．ε-greedy は一定の割合で探索か活用かを選択し，UCB1 は報酬分布の期待値についての信頼区間を用いて行動選択を行うなどの違いがある．こうしたアルゴリズムのメリットとしては，A/B テストのように表示テスト期間を明示的に設定する必要がなく，自動的に有効なコンテンツへと収束していくことが挙げられる．

2.2 Contextual Bandit algorithm

ユーザの特徴量を用いる Contextual bandit algorithm は，これまでに多くのアルゴリズムが提案されている．例えば，UCB1 にコンテキスト情報を加味した LinUCB[10][11][12]，その拡張の BaseLinUCB[13]，LinREL[14]，CoFineUCB[15]や FactorUCB[16]などがある．また，Thompson sampling のコンテキスト拡張[17][18]や，ユーザの潜在クラスを用いた LCB[19]なども提案されている．

2.3 Contextual Bandit algorithm のサービス適用例

Contextual Bandit algorithm はアルゴリズムの研究が中心であるが，いくつかサービス適用事例も報告されている．適用ドメインについても，ランディングページのコンテンツ選択やニュース配信のように，モバイルサービスにおける意思決定によく用いられる．その際，利用するユーザの特徴量は，一般的にモバイルサービスの内容に依存する．

Li らの研究では，Yahoo!のフロントページ上のニュース記事推薦に Contextual bandit algorithm の LinUCB を適用している[10]．また，利用するユーザの特徴量としては，ユーザについての性別や年代といった属性情報や，過去の Yahoo!ページのアクセスログなどがある．この適用によって，特徴量を用いない Context-free bandit algorithm に比べて 12.5%のクリック数増を達成している．

Bouneffouf らの研究では，ユーザへの情報推薦を目的にε-greedy のコンテキスト拡張である Contextual ε-greedy algorithm を用いている[20]．利用しているユーザの特徴量は，ユーザの位置情報や時間，そしてソーシャル情報であり，この 3 種類をオントロジーとして表現している．例えば，対象とするユーザの位置情報として緯度経度が(48.89, 2.23)，該当時刻が”Oct_3_12:10_2012”，ソーシャル情報としては，ユーザのカレンダーに登録してあるイベント情報”meeting with Paul Gerard”を使って，$S =$ ("48.89,2.23", "Oct_3_12:10_2012", "Paul_Gerard")のように表す．

このように，関連研究ではモバイルサービスに依存したサービス利用ログ等を特徴量に用いているが，こうした十分な特徴量が，常に獲得できるとは限らない．また，こうしたデータを新規で取得するためには，ユーザの ID を何らかのサービスと連動する必要があり容易ではない可能性がある．そこで，本研究では，どのようなモバイルサービスでも汎用的に Contextual bandit algorithm を適用できるよう，モバイルネットワークに基づく特徴量のみを利用した手法を提案する．

3. 問題設定

本研究では，モバイルサービスにおけるコンテンツのレコメンデーションを扱う．まず Bandit algorithms を適用する具体的なサービスを述べる．次に，コンテキストの一つモバイルネットワーク特徴量に着目し，その性質を述べる．

3.1 サービス概要

DOCOMO Digital はヨーロッパを拠点とする，グローバル e コマース企業である．その決済プラットフォームは，世界 20 カ国以上の国々に提供し，1 日に 4,200 以上のキャンペーンサイトを運営している．このサイトにおいて，広告バナーをクリックしたユーザにたいし，次に表示するランディングページを出し分けることを考える（図 1）．

Vol.2018-DPS-175 No.19
Vol.2018-MBL-87 No.19
Vol.2018-ITS-73 No.19
2018/5/24

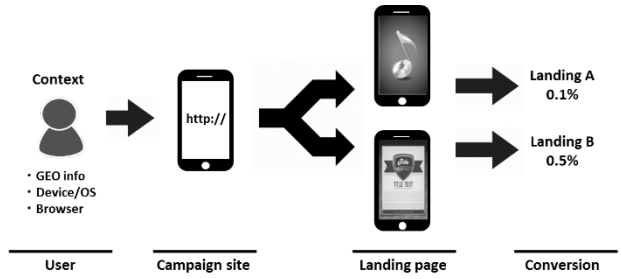

図 1　ランディングページのレコメンデーション
Figure 1　Recommendation of Landing Pages at Campaign Site

ユーザにランディングページを表示した後に，コンバージョンまで達成されたかどうかは，リアルタイムでサーバ側にフィードバックされ，ログに蓄積される．そのため，強化学習のアルゴリズムのひとつである Bandit algorithm を用いての即時的な学習が可能であり，徐々に有効なランディングページへと表示が収束していく．

ここでは，図 1 におけるランディングページでユーザのコンテキストを考慮した上で，効果的なコンテンツを表示し，コンバージョン率を最大化させることを目的とする．

3.2 モバイルネットワーク特徴量

モバイルサービスにおけるコンテンツのレコメンデーションを, Contextual bandit algorithm を用いて実施する際に，ユーザのどのような特徴量が利用できるかは重要な問題である．サービスによっては，ユーザに対する豊富な特徴量が常に取得できているとは限らない．また，そのような多種の特徴量を新たに取得することは，開発コストの増加やプライバシーポリシーの変更等の理由で容易ではない．

一方で，Implicit features であるモバイルネットワーク側の特徴量は獲得コストも低く，加えて，どのようなモバイルサービス上においても共通して利用できる点にメリットがある．本研究で対象とするモバイルネットワークの特徴量の種類の一例を表 1 に示す．特徴量は大きく 5 つに分かれており，①ネットワーク種別②ネットワークオペレーター③ユーザエージェント④オペレーティングシステム⑤ウェブブラウザーが挙げられる．

これらの特徴量は，全てカテゴリカル変数であり，ユーザの特徴量ベクトルは，One-hot-encoding によりバイナリーベクトルとして表現可能である．しかし，これらの特徴量は，非常に多くのカテゴリーが存在するため，一般的に特徴量ベクトルの次元数は膨大になってしまう. そのため，この特徴量ベクトルをそのまま Contextual bandit algorithm の入力としてしまうと，スパース性の問題が発生し，精度が低下するだけでなく, 計算処理にも負荷をかけてしまう．

図 2 では，キャンペーンサイト（Publisher）ごとにモバイルネットワーク特徴量の各カテゴリー（表 1 で示した F00 から F27）についての発生頻度を示す．

表 1　モバイルネットワーク特徴量の例
Table 1　Example of Mobile Network Features

NW 特徴量	内容	No.	カテゴリー値
NW Mode	ユーザがサイトにアクセスした際のネットワーク種別	F00 F01 F02	3G Wi-Fi Unknown
Operator	ユーザが契約しているネットワークオペレーター	F03 F04 F05 F06 F07	Movistar.es Orange.es Vodafone.es Yoigo.es Unknown
User Agent Group	ユーザが利用している移動機についてのユーザエージェント	F08 F09 F10 F11 F12 F13 F14	Android_phone Android_tablet iPhone iPad Windows_smartphone Blackberry Feature_phone
Mobile OS	ユーザが利用している移動機についてのオペレーティングシステム	F15 F16 F17 F18 F19 F20	Android iPhone OS Mac OS X Windows Linux Firefox OS
Mobile Browser	ユーザが利用している移動機のウェブブラウザー	F21 F22 F23 F24 F25 F26 F27	Android Webkit Chrome Mobile Safari Opera Firefox Internet Explore Dofin

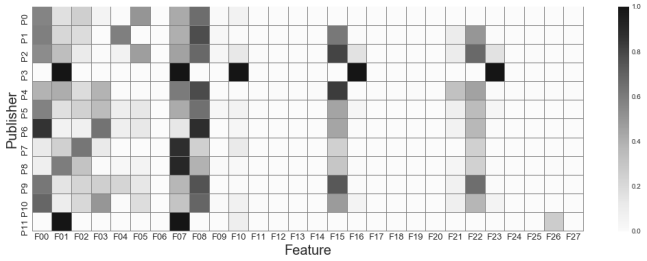

図 2　特徴量のカテゴリー別の頻度
Figure 2　Frequency of Feature Category

ウェブブラウザーやユーザエージェントといったものはカテゴリーの種類が多く存在して，発生頻度が少ないものが多数を占めているスパースな状態になっている．

こうした課題に対処するため，一般的には特徴量を次元圧縮するアプローチがよく用いられる．本論文では，さらに特徴量を扱いやすくするために，次のような処理を行った．

1. 高次元かつスパースなモバイルネットワーク特徴量について, 次元圧縮により連続的な値の特徴量ベクトルに変換し, さらにクラスタリングによって低次元に離散化させる
2. オンラインでのサービスの適用では Contextual-free bandit algorithm と Contextual bandit algorithm を併用し, 動的に次元圧縮とクラスタリングを実行し, その結果に基づいてレコメンデーションを行う

Vol.2018-DPS-175 No.19
Vol.2018-MBL-87 No.19
Vol.2018-ITS-73 No.19
2018/5/24

4. 提案手法

本章では，どのようなモバイルサービスにおいても共通して利用可能なモバイルネットワーク特徴量による Contextual bandit algorithms について解説する．

4.1 提案アルゴリズム

次元圧縮とクラスタリングを実行するためには，ユーザの訪問履歴がある程度，蓄積されてからでないと決定することができない．そのため，次元圧縮とクラスタリングができるまでは特徴量を用いない Context-free bandit algorithm を実行しながら，ユーザについてのデータを獲得していく．そして，一定期間が経った後に，蓄積したデータを基にして次元圧縮とクラスタリングを実行し，コンテキストとして用いるユーザについてのクラスターを作成する．その後の訪問ユーザu_tに対しては，所属クラスターを示すベクトル$\mathrm{Z}_{t,a} \in \{0,1\}^c$を計算した上で，Contextual bandit algorithm を適用する．この Context-free bandit algorithm と Contextual bandit algorithm を併用した提案アルゴリズムを Algorithm1 に示す．

ALGORITHM 1: Context-free and Contextual Bandit

Input: Feature vector X_t of user visited at time t, and the set A_t of advertisement candidates

Output: the selected advertisement $a_t \in A_t$ for the user visited at time t

```
// Context-free bandit phase;
for t = 1, 2, . . . , T_h do
    For each arm i = 1, . . . , K, sample θ_i(t) from the
      Beta(S_i+1,F_i+1) distribution;
    Play arm a(t) := arg max_i θ_i(t);
    Observe reward r_{a,t};
    if r_{a,t} = 1 then
        S_{a,t} = S_{a,t} + 1;
    else
        F_{a,t} = F_{a,t} + 1;
    end
end

// Clustering phase;
do PCA(X_t = {X_1, X_2, . . . , X_{T_h}}, d');
Obtain eigenvectors ξ_j, eigenvalues λ_j and Y_t;
do K-means(Y_t = {Y_1, Y_2, . . . , Y_{T_h}}, c);
Obtain centroids v_c;

// Contextual bandit phase;
for t = T_h + 1, T_h + 2, . . . do
    Observe context X_t of user visited at time t;
    Transform X_t into Z_t ∈ {0, 1}^c by using ξ_j,λ_j,v_c;
    For each arm i = 1, . . . , K, sample μ̃_i(t) from the
      N(μ̂_i(t), v^2 B(t)^{-1}) distribution;
    Play arm a(t) := arg max_i Z_t^T μ̃_i(t);
    Observe reward r_{a,t};
    Update B(t), μ̂_a(t);
end
```

4.2 Context-free bandit phase

本節では Algorithm1 の Context-free bandit phase について説明する．開始当初のフェーズでは，データを蓄積しつつ効果的にコンテンツが表示できるように，コンテキストを考慮しない Context-free bandit algorithm を適用する．本提案手法では，Context-free bandit algorithm として，性能面とリアルタイムでの運用面に優れた Thompson sampling[8]を用いる．Thompson sampling はベイズ戦略の一種で，その腕が最適である事後確率を基にしてランダムに腕を選択する．

時刻tにモバイルサービス上に訪問したユーザをu_t，選択可能なコンテンツを$a_t \in \mathcal{A}_t$とする．またコンテンツの数は$|\mathcal{A}_t| = K$である．時刻$t = 1, 2, \cdots, T$で訪問ユーザu_tに，コンテンツa_tを選択する試行をおこない，そのときの報酬を$r_{a,t}$として学習することを考える．

Thompson sampling のポリシーでは，それぞれの試行において，各コンテンツiの評価値を次のように算出する．コンテンツiを選択した際のコンバージョン成功回数S_iと失敗回数F_iとしたとき，ベータ分布$\mathrm{Beta}(S_i + 1, F_i + 1)$に従う乱数$\theta_i$をランダムに取得する．この操作を$K$個のコンテンツについて繰り返し，式(1)の様に，その値が最も大きいコンテンツa_tを選択する．

$$a(t) \coloneqq \arg\max_i \theta_i(t). \tag{1}$$

そして，選択したコンテンツa_tをユーザに表示した際，コンバージョンが行われたかどうかの結果$r_{a,t}$を観測し，S_iもしくはF_iについての更新をおこなう．

ここで，Context-free bandit algorithm を適用している期間T_hは，パラメータとして与えられ，任意に設定が可能としている．

4.3 Clustering phase

本節では Algorithm1 の Clustering phase について説明する．ユーザu_tの特徴量X_tは，モバイルネットワークに関するカテゴリカルデータをダミー変数化したものを想定している．一般的に，モバイルネットワークにおけるカテゴリデータは多種多様であるため，X_tは高次元かつスパースなバイナリーベクトルになっている．この状態のまま Contextual bandit algorithm の入力とすれば，式（5）における逆行列$B(t)^{-1}$の計算負荷が高くなってしまう．

このような問題を解決するために，図 3 に示すように，次元圧縮とクラスタリングによるアプローチを考える．d次元の特徴量を$X_t \in \{0,1\}^d$について，より低次元のベクトルによって密に表現するために主成分分析を用いる．主成分分析によって得られた低次元d'の特徴量ベクトルを$\mathrm{Y}_t \in \mathbb{R}^{d'}$とする．さらにこれを K-means により，ユーザu_tをクラスター数cでクラスター化させる．最終的に得られた，ユーザがどのクラスターに属しているかを表すベクトル$Z_t \in \{0,1\}^{\mathrm{c}}$を Contextual bandit algorithm の入力とする．

Vol.2018-DPS-175 No.19
Vol.2018-MBL-87 No.19
Vol.2018-ITS-73 No.19
2018/5/24

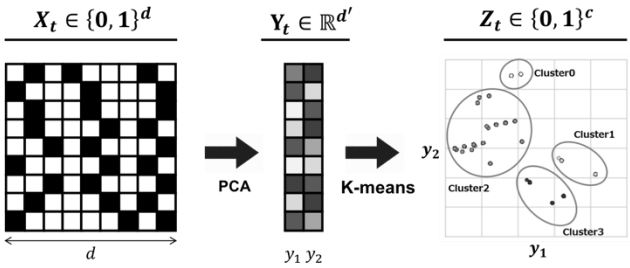

図 3　クラスター作成ステップ
Figure 3　Step of Making Clusters

4.4 Contextual bandit phase

本節ではAlgorithm1のContextual bandit phaseについて説明する．ユーザu_tの特徴量を元のd次元の特徴量$X_t \in \{0,1\}^d$からc次元の特徴量$Z_t \in \{0,1\}^c$に圧縮し，Contextual bandit algorithm を適用する．本提案手法では，Contextual bandit algorithm として Thompson sampling のコンテキスト拡張である[17]を用いる．本アルゴリズムの採用理由は，事前検証にて LinUCB[10]と比較した際に上回っていた為である．

線形モデルで，報酬と特徴量の関係を表現すると，次のようになる．

$$\mathbb{E}[r_{t,a} \mid Z_t] = Z_t^T \mu_a. \tag{2}$$

ここで，$\mu_a \in \mathbb{R}^c$は未知の偏回帰係数である．

各訪問ユーザに対する最適な腕をa_t^*とすると，各試行における最適値と平均報酬との差は次のように表せる．

$$\Delta_t = Z_t^T \mu_{a^*} - Z_t^T \mu_a. \tag{3}$$

従って，目的関数は，全期間Tにおけるリグレット$R(T) = \sum_{t=1}^{T} \Delta_t$の最小化となる．

コンテキスト拡張した Thompson sampling[17]では，ガウシアン尤度関数とガウシアン事前分布を用いる．報酬$r_{a,t}$の尤度，コンテキストX_t，そして偏回帰係数μ_aが正規分布$\mathcal{N}(Z_t^T \mu_a, v^2)$のように与えられるとする．ここで$v = R\sqrt{24/\varepsilon\, k \ln(1/\delta)}$, $\varepsilon \in (0,1), \delta \in (0,1), R \geq 0$である．このとき，$\mu_a$の時刻$t$における推定値は次のように与えられる．

$$B(t) = I_d + \sum_{\tau=1}^{t-1} Z_\tau \cdot Z_\tau^T \tag{4}$$

$$\hat{\mu}_a(t) = B(t)^{-1} \left(\sum_{\tau=1}^{t-1} Z_\tau \cdot r_{a,\tau} \right). \tag{5}$$

得られた推定値$\hat{\mu}_a$を使って，各コンテンツについて正規分布$\mathcal{N}(\hat{\mu}_a(t), v^2 B(t)^{-1})$からサンプリングを実施して，$\tilde{\mu}_a$を得る．そして，式(6)を満たすコンテンツ$a_t$を選択する．

$$a(t) \coloneqq \arg\max_i Z_t^T \tilde{\mu}_i. \tag{6}$$

この選択によって得られた報酬$r_{a,t}$を観測して，$B(t)$および$\hat{\mu}_a(t)$ についての更新をおこなう．

5. オフライン環境でのシミュレーション

本章では，次元圧縮，及びクラスタリングに基づくContextual bandit algorithmの性能検証について述べる．精度検証でのデータは，DOCOMO Digital 社のモバイルサービス上において，過去に実施したコンバージョンについてのA/Bテスト結果を利用している．これらのデータについて次元圧縮とクラスタリングを行い，得られたクラスターをコンテキストとしたオフラインシミュレーション結果について説明する．

5.1 データセット

利用したデータセットは，A.スペイン，及び，B.ギリシャのそれぞれで実施されたキャンペーンサイトにおけるA/Bテスト結果である．どちらのデータセットについても，ランディングページに表示するコンテンツ候補は$K = 2$で，アクセスユーザに対してどちらか一方を選択して表示する．データには，アクセスユーザに関するそれぞれのネットワーク特徴量や，アクセスユーザごとにどのコンテンツを表示したか（インプレッション），及び，表示後にコンテンツを購入したか(コンバージョン)の情報が記録されている．

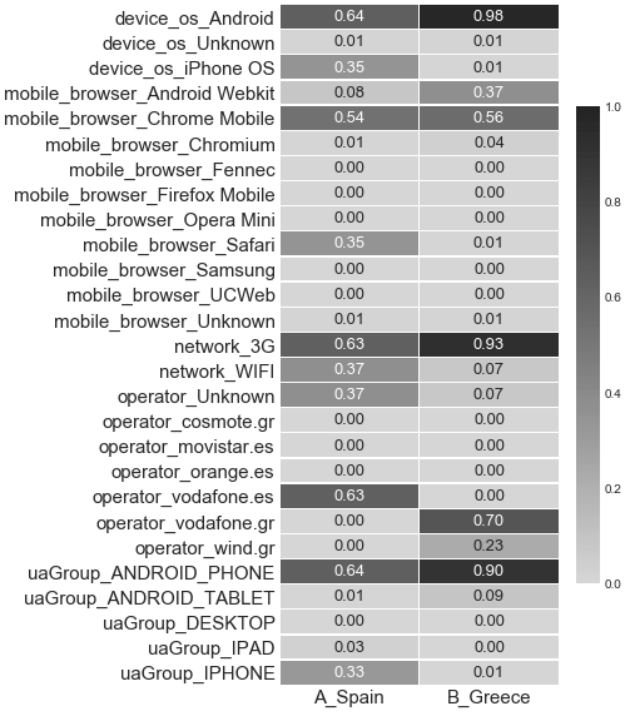

図 4　データセット間の特徴量分布の違い
Figure 4　Difference in Feature Distribution Between Datasets

Vol.2018-DPS-175 No.19
Vol.2018-MBL-87 No.19
Vol.2018-ITS-73 No.19
2018/5/24

図 4 にスペインとギリシャの各キャンペーンサイトでのモバイルネットワーク特徴量の分布を示す．それぞれの特徴量の分布は大きく異なっていることが分かる．例えば，ユーザのオペレーティングシステムについて，スペインでは Android が 64%程度であるが，ギリシャでは 98%となっている．

5.2 シミュレーション設定

シミュレーションにおける報酬は，アクセスユーザに対してコンテンツを表示した際のコンバージョン結果とする．すなわち，ユーザがコンバージョンしていれば$r_{a,t}=1$，コンバージョンしていなければ$r_{a,t}=0$とする．また，アルゴリズムの性能比較については，全アクセスユーザ数Nについてのコンバージョン率である$\sum_t r_{a_{(t)},t}/N$ を評価指標とする．

オフラインでのアルゴリズムシミュレーションをする際，アクセスユーザがコンテンツをコンバージョンするかどうかは，確率変数を使って表現する．この確率変数は，コンテンツiとクラスターkごとのコンバージョン率の実績値$\varphi_{i,k}$をパラメータとするベルヌーイ分布Bernoulli$(x|\varphi_{i,k})$に従うとしている．そして，このN人のアクセスユーザに対する試行を 100 回繰り返すモンテカルロシミュレーションを行い，その時の平均コンバージョン率でアルゴリズムの性能を比較した．

クラスター数cは，以下の理由により$c=4$とした．クラスタリングの前に次元圧縮の目的で行う主成分分析では，本データにおいて，第二主成分までの値で寄与率を概ね占めていた．そして，クラスタリングではその 2 軸に対し，各々の軸の大小で分割させるために$2\times2=4$個のクラスター数とした．

5.3 アルゴリズムの性能比較

2 つのデータセット A, B に対する，モバイルネットワーク特徴量を用いた提案手法でのコンバージョン率を表 2 に示す．アルゴリズムの性能比較のために，特徴量を考慮しない Context-free bandit algorithm，及び，A/B テストで実施した際の結果も示している. 表 2 に示すように, スペイン，及び，ギリシャのどちらのデータセットについても，提案手法が A/B テスト及び Context-free bandit algorithm を上回っていた．特に，データセット A のスペインへのアルゴリズム適用では，A/B テストからは 9%，Context-free bandit algorithm からは 4%のコンバージョン率の向上となっている．これにより，Implicit なコンテキストであるモバイルネットワーク特徴量から作成したユーザについてのクラスターが，Contextual bandit algorithm のコンテキストとして有効であることが分かる．

表 2　アルゴリズムの性能比較

Table 2　Performance Comparison of Algorithms

Algorithms	Dataset	
	A. Spain	B. Greece
Contextual bandit w/ clustering	**4.83%**	**9.92%**
Context-free bandit	4.66%	9.56%
A/B testing	4.43%	8.79%

表 3　クラスター内人数とクラスター特徴

Table 3　Number of People in Clusters and Cluster Description

Cluster	Dataset			
	A. Spain		B. Greece	
	n	description	n	description
Cluster_0	19,253	3G×Androdid	10,098	3G×Operator_A×Browser_A
Cluster_1	4,297	3G×iPhone	5,986	3G×Operator_B
Cluster_2	2,496	Wi-Fi×Android	8,141	3G×Operator_A×Browser_B
Cluster_3	434	Wi-Fi×iPhone	1,976	Wi-Fi

5.4 考察

本節では，モバイルネットワーク特徴量のクラスタリングによって生成されたクラスターの特徴，及び，Contextual bandit algorithm によるコンテンツ選択のクラスター間での違いについて述べ，考察を行う．

5.4.1 クラスターの特徴

オフラインシミュレーションにおいて，4.3 節で述べたように次元圧縮とクラスタリングの操作は，Contextual bandit algorithm の実行前に作成している．その際のクラスター数は$c=4$としており，それぞれのクラスターの特徴を確認した．表 3 に各データセットでの，クラスター内人数とクラスター特徴を示す．どちらのデータセットについても，モバイルネットワークの種別（3G/Wi-Fi）がクラスター形成において重要な因子であることが分かる．さらに，データセット A のスペインでは，ユーザエージェント（Android/iPhone）によって分かれており，データセット B のギリシャでは，通信オレペレーターやモバイルブラウザーの種別によって分かれている．

5.4.2 クラスター別のコンバージョン率比較

コンテキストを考慮しない Context-free bandit algorithm では，全てのユーザに対して同じ期待報酬の分布を仮定しているが，これに対して Contextual bandit algorithm では，それぞれの広告クリエイティブが特徴量に応じた期待報酬の分布を持つと仮定している．そのため，ユーザごとによりマッチしたコンテンツを提供することができるため，Context-fee bandit algorithm に比べて報酬をさらに増やすことができる．本節では、ユーザの特徴量であるクラスターごとにマッチした広告が提供できているか確認する．ここでは，クラスターごとのコンバージョン率を比較する．

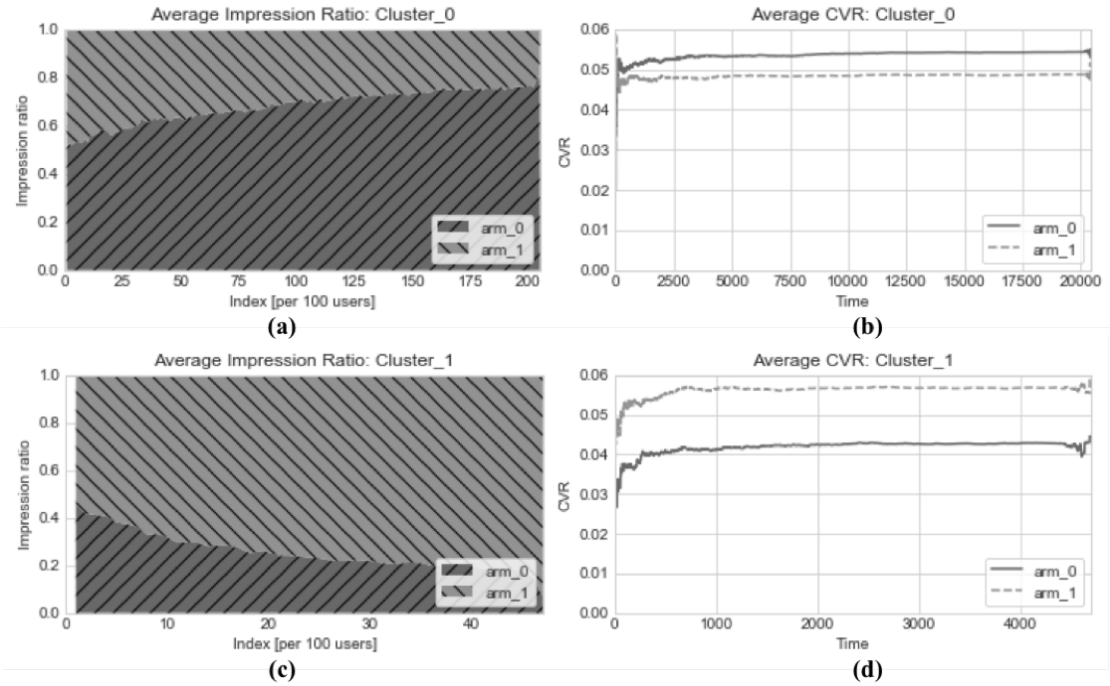

図 5　クラスター別のインプレッション比率（左）とコンバージョン率（右）の推移
Figure 5　Changes in Impression Ratio and Conversion Rate by Cluster
(a) (b) : Cluster_0, (c) (d) : Cluster_1

各データセットにおける 2 つのコンテンツ候補（arm_0, arm_1）のコンバージョン率$\varphi_{i,k}$を表 4 に示す．表中の Total はクラスターを形成しない時のコンバージョン率であり，Diff は arm_0 と arm_1 とのコンバージョン率の差である．データセット A のスペインについて，クラスターを形成しない場合の全体でのコンバージョン率は，$\varphi_0 = 4.73\%, \varphi_1 = 4.50\%$であり，arm_0 のコンバージョン率の方が高く，その差は$\Delta = 0.23\%$である．しかし，クラスターを形成することにより，Cluster_1 では，$\varphi_{0,1} = 4.51\%, \varphi_{1,1} = 5.72\%$となり arm_1 のコンバージョン率の方が 1.21%高くなり，コンテンツ出し分けの効果に繋がったと考えられる．

また，データセット B のギリシャについては，全体で見た場合とクラスター別の場合でも arm_1 のコンバージョン率が支配的である．しかし，arm_0 と arm_1 のコンバージョン率の差の絶対値$|\Delta|$については，Cluster_2, Cluster_3 の差$|\Delta_2| = 2.89\%, |\Delta_3| = 3.49\%$は全体での場合$|\Delta| = 1.60\%$に比べて大きくなっている．これによって context-free bandit と比較して探索の時間が早まり，活用のフェーズへ早期に移行できたために，コンバージョン率が上昇したと考えられる．

5.4.3 Contextual bandit algorithm によるコンテンツ選択

データセット A における，Cluster _0 と Cluster_1 でのインプレッション比率及びコンバージョン率の推移グラフを図 5 に示す．インプレッション比率とは，アクセスユーザを 100 人単位で区切った際に arm_0 及び arm_1 をインプレッションさせた割合を表す．Cluster _0 については，arm_0 のコンバージョン率が高く，アクセスユーザが増えるにつれて，実際に arm_0 のインプレッションを増加できていることが分かる．逆に，Cluster_1 では arm_1 のコンバージョン率の方が高く，arm_1 のインプレッションを増加できていることが分かる．

表 4　コンテンツ・クラスター別のコンバージョン率
Table 4　Conversion Ratio of Content/Cluster Combination

Cluster	Dataset					
	A. Spain			B. Greece		
	arm_0	arm_1	Diff	arm_0	arm_1	Diff
Cluster_0	5.49%	4.94%	0.55%	9.65%	9.95%	-0.30%
Cluster_1	4.51%	5.72%	-1.21%	7.32%	8.28%	-0.96%
Cluster_2	-	-	-	8.70%	11.59%	-2.89%
Cluster_3	-	-	-	0.00%	3.49%	-3.49%
Total	4.73%	4.50%	0.23%	7.99%	9.59%	-1.60%

このように，モバイルネットワーク由来の特徴量を，次元圧縮とクラスタリングによって処理することで，提供するモバイルサービスや国に関係なく有効なコンテキストとして用いることができる．

6. おわりに

本論文では，どのようなモバイルサービス上においても汎用的に Contextual bandit algorithm を適用できるようにするために，モバイルネットワーク由来の特徴量のみを利用する手法について述べた．その特徴量は高次元かつスパー

スであるため，次元圧縮とクラスタリングによってコンテキストを生成した．オフラインでのシミュレーションを行い，スペインのデータセットにおいて，A/B テストからは 9%，Context-free bandit からは 4%のコンバージョン率の向上となった．また，クラスターごとにその特徴やコンバージョン率の差異を検証し，モバイルネットワーク特徴量の優位性を示した．今後の課題としては，ユーザがアクセスしてきた際の位置情報や時間帯をコンテキストとして扱うことにより，さらにユーザにマッチしたコンテンツ推薦が可能になると考えられる．

参考文献

[1] Robbins, H.: Some aspects of the sequential design of experiments, *Bulletin of the American Mathematics Society*, Vol.58, No.5, pp.527–535 (1952).
[2] Auer, P., Cesa-Bianchi, N. and Fischer, P.: Finite-time analysis of the multiarmed bandit problem, *Machine learning*, Vol.47, No.2, pp.235–256 (2002).
[3] Tokic, M.: Adaptive ε-greedy exploration in reinforcement learning based on value differences, *In KI 2010: Advances in Artificial Intelligence*, pp.203–210 (2010).
[4] Cesa-Bianchi, N. and Fischer, P.: Finite-time regret bounds for the multiarmed bandit problem. *Proc. International Conference on Machine Learning*, pp.100–108 (1998).
[5] Auer, P., Cesa-Bianchi, N., Freund, Y., et al.: The nonstochastic multiarmed bandit problem. *SIAM journal on computing*, Vol.32, No.1, pp.48-77 (2002).
[6] Thompson, W.R.: On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples, *Biometrika*, Vol.25, pp.285-294 (1933).
[7] Chapelle, O. and Li, L.: An empirical evaluation of Thompson sampling. *Proc. Advances in neural information processing systems*, pp.2249–2257 (2011).
[8] Agrawal, S., and Goyal, N.: Analysis of Thompson Sampling for the Multi-Armed Bandit Problem, *Proc. Conference on Learning Theory*, p.39.1-39.26 (2012).
[9] Scott, S.L.: A modern Bayesian look at the multi-armed bandit, *Applied Stochastic Models in Business and Industry,* Vol.26, No.6, pp.639-658 (2010).
[10] Li, L., Chu, W., Langford, J., et al: A contextual-bandit approach to personalized news article recommendation, *Proc. The 19th international conference on World wide web*, pp.661-670 (2010).
[11] Li, L., Chu, W., Langford, J., et al: Unbiased Offline Evaluation of Contextual-bandit-based News Article Recommendation Algorithms, Proc. *The fourth ACM international conference on Web search and data mining*, pp.297-306 (2011)
[12] Li, L., Chu, W., Langford, et al.: An unbiased offline evaluation of contextual bandit algorithms with generalized linear models, *Proc. the Workshop on On-line Trading of Exploration and Exploitation 2*, pp. 19-36, (2012).
[13] Chu, W., Li, L., Reyzin, L., et al.: Contextual Bandits with Linear Payoff Functions, *Proc. The Fourteenth International Conference on Artificial Intelligence and Statistics*, pp.208-214 (2011).
[14] Auer, P.: Using Confidence Bounds for Exploitation-Exploration Trade-offs, *Journal of Machine Learning Research*, 3(Nov), pp.397-422 (2002).
[15] Yue, Y., Hong, S.A. and Guestring, C.: Hierarchical Exploration for Accelerating Contextual Bandits. *Proc. The 29th International Conference on Machine Learning*, pp.979-986 (2012).
[16] Wang, H., Wu, Q., and Wang, H.: Factorization Bandits for Interactive Recommendation, *Proc. AAAI Conference on Artificial Intelligence*, (2017).
[17] Agrawal, S. and Goyal, N.: Thompson Sampling for Contextual Bandits with Linear Payoffs. *Proc. The 30th International Conference on Machine Learning*, pp.127-135 (2013).
[18] Lin L.: Generalized Thompson Sampling for Contextual Bandits, *arXiv preprint arXiv:1310.7163*, (2013).
[19] Zhou, L., & Brunskill, E.: Latent contextual bandits and their application to personalized recommendations for new users, *arXiv preprint arXiv:1604.06743*, (2016).
[20] Bouneffouf, D., Bouzeghoub, A., and Gançarski, A. L: A contextual -bandit algorithm for mobile context-aware recommender system, *Proc. International Conference on Neural Information Processing*, pp. 324-331, (2012).
[21] Zhou, L.: A Survey on Contextual Multi-armed Bandits, *arXiv preprint arXiv:1508.03326*, (2015).
[22] “Apache Spark”, https://spark.apache.org/