情報処理学会論文誌 Vol.60 No.1 38–47 (Jan. 2019)

モバイルネットワーク特徴量のクラスタリングによる Contextual Bandit Algorithm

出水 宰[1,a)] ルーベン マンサーノ[2] セルジオ ゴメス[2] 深澤 佑介[1]

受付日 2018年4月11日, 採録日 2018年10月2日

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つね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で行った検証で，スペインでは提案手法によって A/B テストに比べて 9.0%，特徴量を考慮しない Bandit algorithm に比べて 3.6%のコンバージョン率の向上を確認した．また，アドプラットフォームで実施したスペインでのオンラインテストでは，提案手法によって A/B テストに比べて 15.6%のコンバージョン率の向上を確認した．

キーワード：バンディットアルゴリズム，意思決定，広告配信，モバイルネットワーク

A Contextual Bandit Algorithm Based on Clustering of Mobile Network Features

Tsukasa Demizu[1,a)] Rubén Manzano[2] Sergio Gómez[2] Yusuke Fukazawa[1]

Received: April 11, 2018, Accepted: October 2, 2018

Abstract: As a method for increasing advertisement effect on a mobile service, the contextual bandit algorithm that selects the content in consideration of the features of the access user is often used. However, it is not always that we can acquire sufficient features of the user. In this paper, we propose a dynamic approach based on dimensional reduction and clustering by using only the features of mobile network information and aim to be able to apply contextual bandit algorithm for any mobile service. We apply the proposed method to the datasets in the campaign sites in Spain and Greece, and it improved the conversion rate in Spain by 9.0% compared to the A/B testing and by 3.6% compared to the bandit algorithm without considering the context. In the on-line test conducted by applying the proposed method to our Ad platform, the proposed method improved the conversion rate in Spain by 15.6% compared to the A/B testing.

Keywords: multi-armed bandit algorithm, decision-making, advertisement recommendation, mobile network

1. はじめに

Web 上での広告配信において，効果の高いコンテンツを選択することは重要である．広告クリエイティブの効果として，広告が表示された回数（インプレッション数）にたいするクリックされた回数の割合であるクリック率や，広告がクリックされた回数にたいする掲載商品が購買された回数（コンバージョン数）の割合であるコンバージョン率といった指標が用いられる．たと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，クリック率やコンバージョン率は変わってくる．しかし，意思決定者は，それら

[1] 株式会社 NTT ドコモ
NTT DOCOMO, INC., Chiyoda, Tokyo 100–6150, Japan
[2] DOCOMO Digital Limited, Calle del Príncipe de Vergara, Madrid, Spain
[a)] tsukasa.demizu.sp@nttdocomo.com

の広告クリエイティブの効果を事前に知ることはできない．そのため，複数ある広告クリエイティブにユーザをランダムに振り分けて，効果を検証する Web マーケティング手法の A/B テストがよく用いられる．

意思決定者が，広告クリエイティブを選択する際に，大きく 2 つの戦略をとることが可能である．1 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，exploration 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，すべ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-free 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，獲得コストは十分に低い．

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）がサービス側でつね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．本論文の貢献内容は，次のとおりである．

- 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．
- 実サービスへの適用で，提案手法の優位性を確認した．

本論文の構成は以下のようである．2 章では，Contextual bandit algorithm に関連する研究について紹介する．3 章では，対象のサービスや問題設定について定義する．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 algorithms について述べる．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．6 章では，オンライン環境での適用結果を説明する．

2. 関連研究

本章では，Bandit algorithm，Contextual bandit algorithm および，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，exploration 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あげられる．

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．たとえば，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 種類をオントロジとして表現している．たとえば，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などを特徴量に用いているが，こうした十分な特徴量が，つね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．

図 1 ランディングページのレコメンデーション
Fig. 1 Recommendation of landing pages at campaign site.

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 1 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**）．

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 1 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うえで，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つね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ポリシの変更など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

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	3G
		F01	Wi-Fi
		F02	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	Movistar.es
		F04	Orange.es
		F05	Vodafone.es
		F06	Yoigo.es
		F07	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	Android_phone
		F09	Android_tablet
		F10	iPhone
		F11	iPad
		F12	Windows_smartphone
		F13	Blackberry
		F14	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	Android
		F16	iPhone OS
		F17	Mac OS X
		F18	Windows
		F19	Linux
		F20	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	Android Webkit
		F22	Chrome Mobile
		F23	Safari
		F24	Opera
		F25	Firefox
		F26	Internet Explore
		F27	Dofin

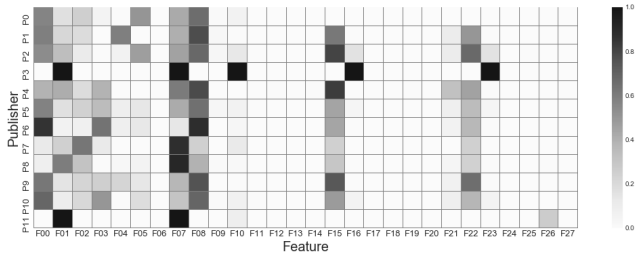

図 2 特徴量のカテゴリ別の頻度
Fig. 2 Frequency of feature category.

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，③ユーザエージェント，④オペレーティングシステム，⑤ウェブブラウザがあげられる．

これらの特徴量は，すべ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（表 1 で示した F00 から F27）についての発生頻度を示す．ウェブブラウザやユーザエージェントといったものはカテゴリ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 Context-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．

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．ここで，Context-free bandit algorithm を適用している期間を T_h と表す．開始から期間 T_h 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，所属クラスタを示すベクトル $\mathrm{Z}_{t,a} \in \{0,1\}^c$ を計算したうえで，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 Algorithm 1 に示す．

4.2 Context-free Bandit Phase

本節では Algorithm 1 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 $t = 1, 2, \ldots, T$ で訪問ユーザ u_t に，コンテンツ a_t を選択する試行を行い，そのときの

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 ν_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,ν_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
```

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 θ_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) のように，その値が最も大きいコンテンツ a_t を選択する．

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

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 についての更新を行う．

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る．適用するモバイルサービスにおけるユーザのアクセス頻度や，扱うべきユーザの特徴量の数に応じて適切な値 T_h は変わってくると考えられるため，ここでは，意思決定者が任意に設定できる値としている．

4.3 Clustering Phase

本節では Algorithm 1 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．

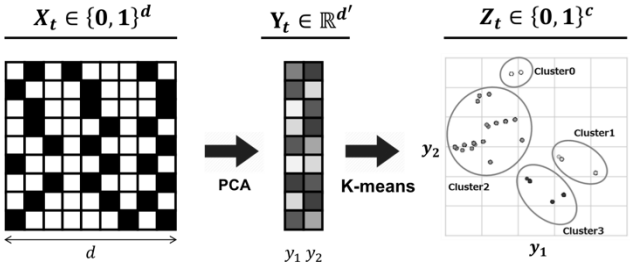

図 3 クラスタ作成ステップ
Fig. 3 Step of making clusters.

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（PCA）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 をクラスタ数 c でクラスタ化させる．最終的に得られた，ユーザがどのクラスタ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．

4.4 Contextual Bandit Phase

本節では Algorithm 1 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文献 [17] を用いる．本アルゴリズムの採用理由は，事前検証で LinUCB [10] と比較した際に上回っていたためである．

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．

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

ここで，$\mu_a \in \mathbb{R}^{\mathrm{c}}$ 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．

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

したがって，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，μ_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 $\mathcal{N}(\hat{\mu}_a(t), v^2B(t)^{-1})$ 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．

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

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 $\hat{\mu}_a(t)$ についての更新を行う．

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，およ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，および，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．

データセット A，B のどちらについても，ランディングページに表示するコンテンツ候補の種類は $K = 2$ であり，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．Bandit algorithm の文脈では，選択肢のことを腕（arm）と呼ぶ．そのため，今回のデータセットにおいても，$K = 2$ のコンテンツ候補のそれぞれを，arm_0，arm_1 と定義する．データセット A と B におけるコンテンツ候補（arm_0，arm_1）はそれぞれ異なるものである．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ちらのコンテンツ候補 (arm_0，arm_1) を表示したか（インプレッション），および，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 にスペインとギリシャの各キャンペーン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．たとえば，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．

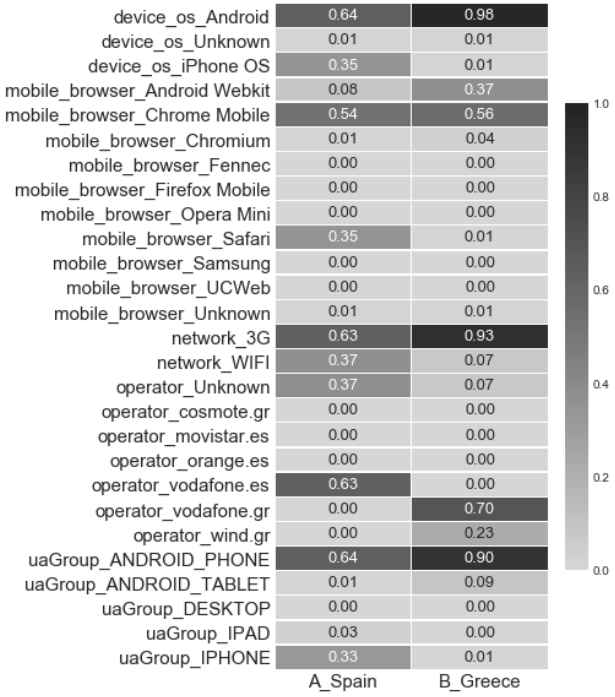

図 4 データセット間の特徴量分布の違い
Fig. 4 Difference in feature distribution between datasets.

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の結果と定義する．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 におけるコンバージョン率である $\sum_t r_{a_{(t)},t}/N$ を評価指標と定義する．

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．本シミュレーションでは，一般的によく用いられる Bernoulli bandit problem [8] を想定する．すなわち，コンテンツ i の報酬は 0 もしくは 1 で表され，報酬が 1 である確率はパラメータ φ_i のベルヌーイ分布 Bernoulli$(x \mid \varphi_i)$ に従うとする設定である．提案手法では，クラスタごとに異なる報酬分布を仮定しているため，ベルヌーイ分布のパラメータ φ_i もクラスタによって異なる．本シミュレーションでは，コンテンツ i とクラスタ k ごとのコンバージョン率の実績値 $\hat{\varphi}_{i,k}$ をパラメータとして利用する．すなわち，クラスタ k のアクセスユーザにコンテンツ i を表示した際のコンバージョン結果は，ベルヌーイ分布 Bernoulli$(x \mid \hat{\varphi}_{i,k})$ に従う．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

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 クラスタ内人数とクラスタ特徴
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×Android	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

カルロシミュレーションを行い，そのとき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 2 主成分までの値で寄与率をおおむね占めていた．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 $2 \times 2 = 4$ 個のクラスタ数とした．

本シミュレーションでは，モバイルネットワーク特徴量を利用した Contextual bandit algorithms の優位性を検証することを目的としている．そのため，データセット A および B が得られている状態で，特徴量についての次元圧縮とクラスタリングを実行し，Contextual bandit algorithms を適用した．したがって，本シミュレーションでは Context-free bandit から Clustering phase に移行するまでの期間 T_h は含まれていない．

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，および，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，および，ギリシャのどちらのデータセットについても，提案手法によるコンバージョン率が A/B テスト，および，Context-free bandit algorithm によるコンバージョン率を上回っていた．

ここで，手法 i でのコンバージョン率を r_i として，手法 i から手法 j へ変えたときのコンバージョン率の向上率 $R_{i,j}$ を $R_{i,j} = (r_j - r_i)/r_i$ と定義する．データセット A について，A/B テストから提案手法への向上率は 9.0%，Context-free bandit algorithm から提案手法への向上率は 3.6%となっている．同様に，データセット B について，A/B テストから提案手法への向上率は 12.9%，Context-free bandit algorithm から提案手法への向上率は 3.8%となっている．提案手法による，これら 4 つの向上率について，t 検定を行い，すべてにおいて有意差が認められた（$p < 0.05$）．

これらの結果により，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．

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の特徴，および，Contextual bandit algorithm によるコンテンツ選択のクラスタ間での違いについて述べ，考察を行う．

5.4.1 クラスタの特徴

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．その際のクラスタ数は $c = 4$ としており，それぞれのクラスタの特徴を確認した．**表 3** に各データセットでの，クラスタ内人数とクラスタ特徴を示す．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，通信オペレータやモバイルブラウザの種別によって分かれている．

5.4.2 クラスタ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，すべ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-free bandit algorithm に比べて報酬をさらに増やすことができる．本項では，ユーザの特徴量であるクラスタ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，クラスタごとのコンバージョン率を比較する．

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）のクラスタ別のコンバージョン率 $\varphi_{i,k}$ を**表 4** に示す．表中の Total はクラスタを形成しないとき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．

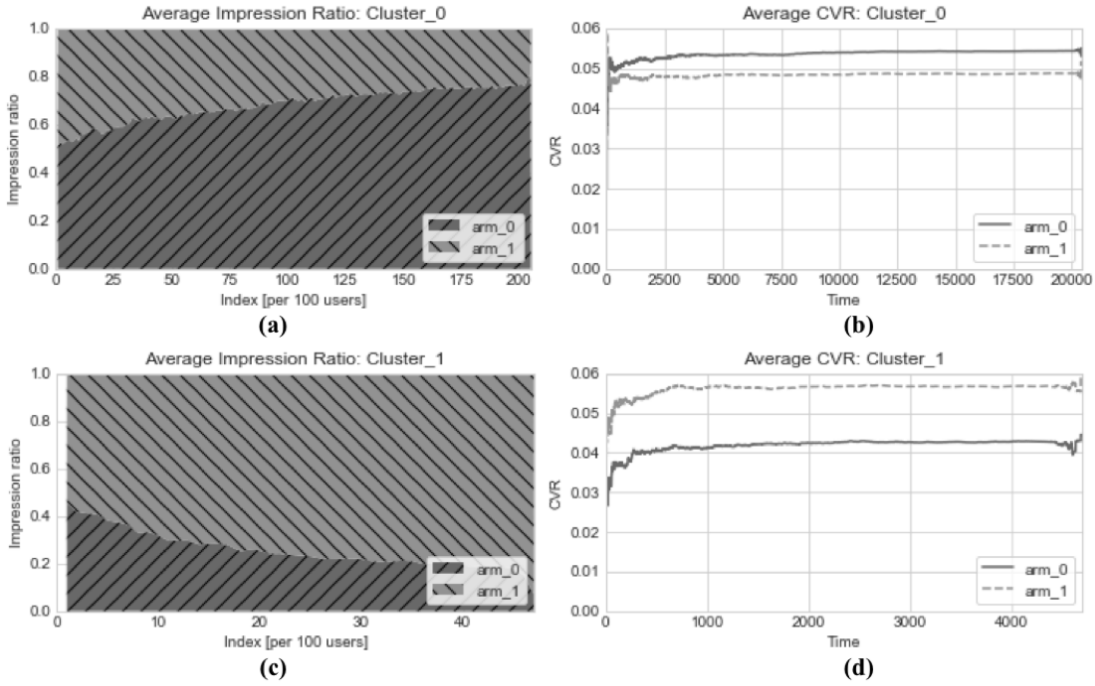

図 5 クラスタ別のインプレッション比率（左）とコンバージョン率（右）の推移
Fig. 5 Changes in impression ratio and conversion rate by cluster. (a) (b): Cluster_0, (c) (d): Cluster_1.

表 4 コンテンツ・クラスタ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%

しかし，クラスタ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つな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およ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 および 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させることが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させることができていることが分かる．

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．

なお，本シミュレーションではコンテンツ候補数を $K = 2$ としているが，一般的には，K が 2 よりも十分に大きい状況も考えられる．しかし，コンテンツを選択する際の判断基準である式 (1) および (6) より，十分なアクセスユーザを得ることでアルゴリズムによる選択は収束することから，K の値によるアルゴリズムへの性能影響はないと考えられる．

6. オンラインでのサービス適用

提案手法である Algorithm1 のオンラインサービスへの適用には，**図 6** で示すように，期間 T_h までは Context-free bandit を使ってデータを蓄積した後に，次元圧縮とクラスタリングを実行する．そして，期間 T_h 以降は，アクセスユーザに対して所属するクラスタをコンテキストとして Contextual bandit を適用する．また，次元圧縮とクラスタリングについては，オンライン上で高速に実施できるよ

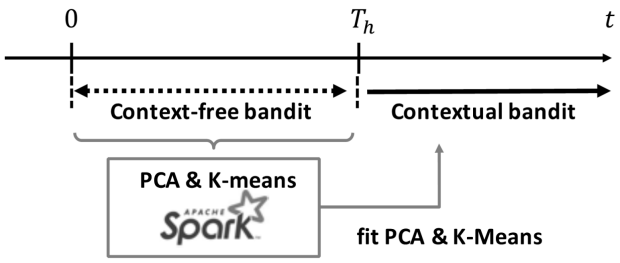

図 6 オンラインサービスへの適用
Fig. 6 Application to on-line service.

表 5 データ概要
Table 5 Data summary.

Variable	Value
Data period	From 28-08-2017 To 17-09-2017
Country	Spain
T_h	72 hours
K	2
c	4

う，Apache Spark [22] を用いている．

このアーキテクチャを DOCOMO Digital 社のアドプラットフォーム上に構築し，提案アルゴリズムでのオンラインテストを実施した．**表 5** に実施したオンラインテストでのデータ概要を示す．検証はスペインのキャンペーンサイトで行い，期間は 2017 年 8 月 28 日から 9 月 17 日までの 3 週間で適用した．また Context-free bandit の適用期間 T_h は 72 時間に設定した．ここでパラメータである期間 T_h の決定方法については，5 章で述べたオフライン検証で利用したデータセット A および B のデータ期間と同じ時間である 72 時間を採用している．コンテンツ候補数 K とクラスタ数 c は，オフライン検証時と同じ $K = 2$，$c = 4$ としている．この設定のもとで，A/B テストでの結果と比較して，15.6%のコンバージョン率向上を確認した．

7. 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では，スペインのデータセットにおいてコンバージョン率の向上率が，A/B テストから提案手法へは 9.0%，Context-free bandit algorithm から提案手法へは 3.6%であった．また，クラスタ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．オンラインへのサービス適用については，Context-free bandit と Contextual bandit とをハイブリッドした手法を提案し，アーキテクチャを構築した．スペインのキャンペーンサイトで行ったオンラインテストでは，コンバージョン率の向上率が，A/B テストから提案手法へは 15.6%を確認した．

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．また，データセットによっては，コンテンツごとのコンバージョン行動と特徴量とに相関がほとんど存在しない場合も考えられる．このように相関の度合いによる提案アルゴリズムへの影響分析も今後の課題としてあげられる．

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, *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. 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. 4th 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. 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. 14th 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. 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. 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. and 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, available from ⟨https://spark.apache.org/⟩.

出水 宰

2011 年大阪大学工学部卒業．2013 年大阪大学大学院情報科学研究科博士前期課程修了．同年株式会社 NTT ドコモ入社．2018 年 10 月より大阪大学大学院情報科学研究科博士後期課程に進学．マーケティングにおける予測モデル，数理最適化，強化学習に関する研究開発を行っている．日本オペレーションズ・リサーチ学会会員．

ルーベン マンサーノ

2002 年 DOCOMO Digital 入社．デジタルマーケティングに関するサービス開発を行っている．2014 年 IESE Business School (University of Navarra) にて General Management Program を修了．

セルジオ ゴメス

2008 年 DOCOMO Digital 入社．アドテクノロジ，決済サービス，異常検知に関する開発を行っている．Carlos III university にて Master of Computer Science を修了．

深澤 佑介 （正会員）

2002 年東京大学工学部卒業．2004 年東京大学大学院工学系研究科修士課程修了．同年株式会社 NTT ドコモ入社．2011 年東京大学大学院工学系研究科博士後期課程修了．同年 10 月より東京大学人工物工学研究センターにて協力研究員，2017 年より客員研究員兼任，現在に至る．Web マイニング，レコメンデーション，実世界行動予測に関する研究開発を行っている．IEEE，人工知能学会各会員．博士（工学）．