

# ユーザのシチュエーションと振る舞いに基づく 提示タイミング制御を用いたニュースリーダー

樽見 彰仁<sup>†</sup> 北山 大輔<sup>†</sup>

<sup>†</sup> 工学院大学 大学院工学研究科 情報学専攻 〒163-8677 東京都新宿区西新宿 1-24-2

E-mail: <sup>†</sup>tem19010@ns.kogakuin.ac.jp, <sup>††</sup>kitayama@cc.kogakuin.ac.jp

あらまし Web クリップアプリ、ニュースフィードアプリには、コンテンツ一覧を表示したり、コンテンツを保存できる機能がある。そのコンテンツの中で時間の経ったもの、保存し忘れたものや、その時のユーザがコンテンツを読むことができない状況である場合、コンテンツは埋もれたままになり、再びそのコンテンツにアクセスすることは容易ではない。本研究では、この問題を解決するために、過去に閲覧した、ユーザの興味があるようなコンテンツとして、そのコンテンツの内容とシチュエーション（時間、場所）を用いて学習し、ユーザモデルを作成する。作成したユーザモデルを用いて、ユーザの現在のシチュエーションにおける、コンテンツの適切さを判定し、適切である場合に、過去に見落としていたであろう未閲覧のコンテンツを推薦するシステムを構築し、評価を行う。

キーワード ニュース推薦、推薦システム、SVR、振る舞い、シチュエーション

## 1 はじめに

私たちは、記憶を補助する目的で、コンテンツやタスクを保存するアプリを利用している。コンテンツを保存するためのアプリとして、ユーザが後で閲覧したい Web ページを保存する Pocket<sup>1</sup>などの Web クリップアプリや、ユーザがフォローしたニュースサイトの新着記事を取得・保存が可能な Feedly<sup>2</sup>などがある。また、タスク管理をするためのアプリとして、期日になったときタスクをリマインドする Todoist<sup>3</sup>、Trello<sup>4</sup>などがある。これらの Web クリップアプリ、ニュースフィードアプリ、リマインダアプリには、コンテンツの一覧を表示する機能、コンテンツを保存する機能がある。しかし、これらのアプリには、ユーザが見落とししたり、保存し忘れたりしたコンテンツや、その状況で閲覧することが出来なかったコンテンツなどに再びアクセスすることは容易ではないという問題がある。本研究の目的は、ユーザが見落したコンテンツや、保存したこと自体を忘れたコンテンツを、ユーザに適切なタイミングで推薦し、ユーザの知識獲得を支援して、この問題を解決することである。

そこで本研究では、ユーザが見落とししたり、保存したこと自体を忘れたコンテンツを、ユーザに適切なタイミングで推薦することによって、ユーザの知識獲得を支援することは可能であるかを研究課題とする。ユーザに推薦する適切なタイミングは、過去に閲覧したユーザが興味があるようなコンテンツとして、そのコンテンツの内容とシチュエーション（時間、場所）を用いて学習し、ユーザモデルを作成する。作成したユーザモデルを用いて、ユーザの現在のシチュエーションにおける、コン

テンツの適切さを判定し、適切である場合に、過去に見落としていたであろう未閲覧のコンテンツを推薦する。本研究では、コンテンツとして、ニュース記事を取り扱う。

2 節では、ユーザモデルの作成、ニュース推薦についての関連研究について述べる。3 節では、提案手法について述べる。4 節では、モデルの評価実験について述べる。5 節では、ユーザによる評価実験について述べる。6 節では、まとめと今後の課題について述べる。

## 2 関連研究

ユーザの行動や、位置情報などのコンテキストを用いたモデル作成や、作成したモデルを用いた推薦システムについての研究が行われている。土岐ら [1] は、SNS のストリーム閲覧時のユーザの行動から、ユーザプロフィールを構成する手法の提案した。スクロール動作からツイートの滞留時間を推定したものと、TF-IDF の組み合わせより興味単語を抽出し、TF-IDF に比べ高い精度でプロフィールを構成できることを示した。Liu ら [2] は、ニュース推薦とアクティブタイム予測を組み合わせた DNN モデルを提案した。ユーザのアクティブタイムを予測することによって既存の手法よりも高い精度で推薦が行えることをニュース推薦のためのデータセット Adressa [3] を用いた検証で示した。小林ら [4] は、Twitter 上の噂の元ツイートを RT したユーザのプロフィールと過去ツイートを時系列化した伝播経路を用いて、早期検出を行った。プロフィールと過去ツイート両方を学習することによってプロフィールのみの従来手法 [5] より高い精度で噂を検出できることを示した。奥ら [6] は、ユーザの時系列的な行動パターンがその時々のユーザコンテキストによって変化することを考慮にした、情報推薦方式を提案した。中村ら [7] は、ユーザの複数の興味・関心が混在している Web 上の行動履歴を、事前に分類の基準を策定せず分類し、

1 : <https://getpocket.com/>

2 : <https://feedly.com>

3 : <https://todoist.com/>

4 : <https://trello.com/>

より適応範囲の広い超個人特化プロフィールを生成するための分析を行った。各要素の比率を決めるためさまざまな距離関数を作成し、クラスタリング結果を分析した。その結果、閲覧時刻とリファラーに関しては均等にするほうが適していることを示した。Marios ら [8] は、ユーザのインタラクション、位置情報から、ルール [9] を作成し、ユーザがどのようにニュースを閲覧するのか、どのようにニュース記事を選択するのか、どこで記事を閲覧するかモデル化を行ったが、ルールベースの提案手法より、RandomForest を用いた手法が優れていることを示した。

また、位置情報や加速度センサなどを用いて、屋内環境でユーザの位置を推定する研究が行われている。Azamat ら [10] は、屋内環境での、ユーザの行動と場所のコンテキストからユーザのプロフィールを推定するフレームワークの提案した。ユーザの行動は腕につけたスマートフォンの加速度センサーデータを DeepSense [11] を用いて分類し、場所についてのコンテキストは予め屋内マップを作成し GPS と WiFi から判定し、限られた環境で高い精度で分類が可能であることを示した。Watanabe ら [12] は、止まっている、歩いている、車に乗っているなどのスマートフォンユーザの行動を位置情報から分類を行い、高い精度で分類が可能であることを示した。Ashbrook ら [13] は、シングルユーザ、マルチユーザ両方で、重要な場所を、ユーザ行動した GPS データから取得する手法を提案した。笹井ら [14] は、滞在場所の検出、行動にタスクやコメントを付与できるライフログ FLAG を提案した。滞在場所の検出についての有用性を実験で示した。

本研究は、ニュース推薦のためのユーザモデルの作成において、実空間上のユーザの記事を閲覧した場所と、現時点でのユーザのいる場所を考慮している点が従来手法と異なる。また、コンテンツを閲覧している場所をセンサーデータなどを利用せずに、ユーザのシチュエーションとして学習させる点が従来手法と異なる。

### 3 提案手法

提案手法はユーザの行動ログの収集、ユーザモデルの作成、コンテンツ推薦から成る。はじめに、プロトタイプシステムを用いて、ユーザの閲覧行動をログとして保存する。次に、取得したログデータから、目的変数をユーザが過去にコンテンツを閲覧したかどうか、説明変数をコンテンツ内容、ユーザのシチュエーションとして、SVR を用いて、ユーザモデルを作成する。最後に、作成したユーザモデルを用いて、ユーザの現在閲覧しているコンテンツとシチュエーションから、過去に取得したことのあるコンテンツをユーザに推薦する。提案手法の概要を図 1 に示す。

#### 3.1 ログデータ収集

ログデータ収集のためのアプリケーションについてはプロトタイプシステムを用いる。プロトタイプシステムではユーザの行動を取得し、データベースに保存することができる。ログと

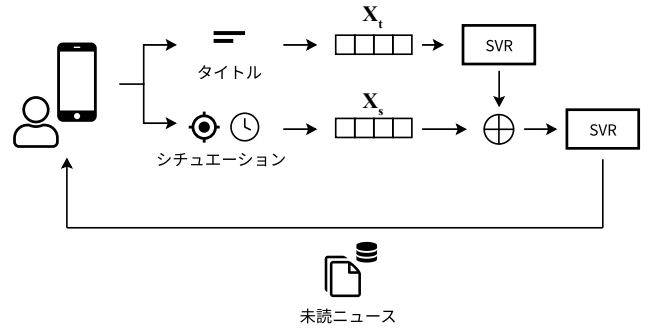


図 1 提案手法

して保存するデータの中で、モデル作成に利用するデータを表 1 に示す。

#### 3.2 目的変数

目的変数  $\mathbf{Y}$  (式 1) は、ユーザがコンテンツを閲覧する可能性として、ユーザが過去にコンテンツ  $C_i$  を閲覧したかどうか  $y_i$  を用いる。  $y_i$  は、  $C_i$  のログデータから、記事の要約文を開いた場合 0.5、記事の全文を開いた場合 1.0、それ以外の場合 0.0 とする (式 2)。なお、後述するプロトタイプの仕様により、要約文はアプリ内で開くことができるが、全文は別途ブラウザが立ち上がりそこで閲覧を行う。  $n$  はユーザが過去に取得した記事の総数である。

$$\mathbf{Y} = [y_0, y_1, \dots, y_n] \quad (1)$$

$$y_i = \begin{cases} 0.5 & \text{if isOpened} \\ 1.0 & \text{if isOpenedInBrowser} \\ 0.0 & \text{Otherwise} \end{cases} \quad (2)$$

#### 3.3 説明変数

##### 3.3.1 コンテンツ内容

コンテンツ内容の説明変数  $\mathbf{X}_t$  として、ニュース記事のタイトルを 300 次元で分散表現したものを用いる。分散表現のモデルは、2020 年 11 月 02 日 02:57 時点での、Wikipedia のダンプデータ<sup>5</sup>を MeCab [15] (辞書: mecab-ipadic-NEologd 2020-08-20) で分かち書きしたものを fastText [16] で 300 次元で学習したものを用いた。作成した分散表現モデルを用いて、fastText の `get_sentence_vector()` 関数にニュース記事のタイトルを入力し、コンテンツ内容の説明変数として利用する。

##### 3.3.2 ユーザのシチュエーション

ユーザのシチュエーションの説明変数  $\mathbf{X}_s$  として、ユーザがニュース記事を閲覧した時刻  $\sin(\pi t), \cos(\pi t)$ 、場所  $x_{l_0}, \dots, x_{l_3}$  を用いる (式 3)。  $\sin(\pi t), \cos(\pi t)$  はログデータの “createdAt” の時刻  $H$ , 分  $M$ , 秒  $S$  から 24 時間を円周上の座標として表現したものである。  $x_{l_0}, \dots, x_{l_3}$  は、ログデータの “location” をダミー変数化したものである、

5: <https://ja.wikipedia.org/wiki/Wikipedia:データベースダウンロード>

表 1 モデル作成に使用するログデータ

キー	型	説明
title	String	ニュース記事のタイトル
location	String	閲覧した場所, ユーザが手動で入力する (private, home, work, public)
isOpened	Bool	要約文を開いたか
isOpenedInBrowser	Bool	全文を開いたか
createdAt	Datetime	ニュースの取得時間

$$\mathbf{X}_s = [\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_n] \quad (3)$$

$$\mathbf{x}_i = [\sin(\pi t), \cos(\pi t), x_{l_0}, \dots, x_{l_3}] \quad (4)$$

$$t = (H \times 3600 + M \times 60 + S) / 86400 \quad (5)$$

### 3.4 回帰モデルの学習

非線形回帰手法として Suport Vector Regression(SVR) を用いる (式 6).  $K(\mathbf{x}_i, \mathbf{x})$  はカーネル関数である.

$$f(\mathbf{x}) = \sum_{i=0}^n (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b$$

$$b = y_i - \sum_{i=0}^n (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + \epsilon \quad (6)$$

$$\sum_{i=0}^n (\alpha_i - \alpha_i^*) = 0 \text{ and } \alpha_i, \alpha_i^* \in [0, C]$$

SVR の実装は, scikit-learn<sup>6</sup>の SVR() クラスを利用した.

### 3.5 実 装

提案手法を実装したアプリケーションについて解説する. アプリケーションの画面を図 2 に示す. アプリケーションを起動すると場所を入力するモーダルが表示される. 場所のボタンをタップするとホーム画面へ遷移する. ホーム画面では, Feedly API から取得した各ニュースタイトルと, ユーザの現在地を変更できるボタンが表示される. ニュースタイトルをタップすると, ニュースのタイトル, ニュースサイトの URL, 公開日時, 要約文が表示され, ニュース記事をブラウザで開く, “続きを読む” ボタンと, 記事を保存できる “保存する” ボタンが表示される. ここで, ニュース記事が推薦記事である場合, 公開日時の下に評価ボタンが表示される. ユーザはこの評価ボタンで推薦が適切であるかどうかを評価することができる. 推薦記事は, Feedly から取得した記事 3 件ごとに挿入し, 最大 5 件挿入される.

また, アプリケーションは, Web, Android, iOS 上で動作するよう Flutter<sup>7</sup>でアプリケーションを作成した. ログ (表 1) はユーザアカウントごとに Firebase<sup>8</sup>の Firestore データベースに保存される. また, このアプリケーションを利用する際は, ユーザは事前に Feedly API と連携する必要がある.

## 4 モデル評価実験

説明変数の組み合わせごとに 7 つの手法を比較する. 学習

データとテストデータとして, 著者自身がアプリケーション (図 2) を利用して収集したデータ (ニュース 472 件分) を利用した. 学習データのテストデータはそれぞれ 75% と 25% の割合でランダムに分割し. 平均 0, 分散 1 で標準化したものを利用した. パラメータについては, GridSearchCV() による交差検証でパラメータの探索を行った. 分割数は 3, カーネルは rbf (ガウスカーネル), poly (多項式カーネル)  $C, \epsilon$  については  $[0.001, 0.01, 0.1, 1]$  の範囲で値で探索した. 評価指数として決定係数  $R^2$  を利用した.

### 4.1 学 習 結 果

説明変数の組み合わせと, 学習時の交差検証の結果, 最も優れていたパラメータを表 2 に示す. 1. は説明変数として, タイトルから得られた  $\mathbf{X}_t$  のみを用いた. 2. はユーザのシチュエーション  $\mathbf{X}_s$  のうち, 閲覧時の時間のみの  $\mathbf{X}_{st} = [\sin(\pi t), \cos(\pi t)]$  を用いた. 3. はユーザのシチュエーション  $\mathbf{X}_s$  のみを用いた. 4. はタイトル  $\mathbf{X}_t$  と時間のみのユーザのシチュエーション  $\mathbf{X}_{st}$  を結合したものをを用いた. 5. はタイトル  $\mathbf{X}_t$  とユーザのシチュエーション  $\mathbf{X}_s$  を結合したものをを用いた. 6. は 1. の出力結果と時間のみのユーザのシチュエーション  $\mathbf{X}_{st}$  を結合したものをを用いた. 提案手法である 7. は 1. の出力結果とユーザのシチュエーション  $\mathbf{X}_s$  を結合したものをを用いた. 決定係数  $R^2$  を見ると  $f(\mathbf{X}_t) \oplus \mathbf{X}_s$  が最も高い値となり, 次点で  $f(\mathbf{X}_t) \oplus \mathbf{X}_{st}$ ,  $\mathbf{X}_s$  と続いている.  $\mathbf{X}_t$ ,  $\mathbf{X}_{st}$ ,  $\mathbf{X}_t \oplus \mathbf{X}_{st}$ ,  $\mathbf{X}_t \oplus \mathbf{X}_s$  については, 決定係数  $R^2$  が低く, 学習が失敗している可能性がある.

### 4.2 評 価 結 果

手法ごとの学習モデルにテストデータを入力し, 正解ラベルと比較した結果を表 3 に示す. 青色の Label が正解データ, オレンジ色の Predicted が予測値である. 学習の結果決定係数が特に低かった  $\mathbf{X}_{st}$  は, 予測が不可能であった.  $\mathbf{X}_s$  は, 高い精度で予測できなかったものの, ユーザが読むであろうと予測した 5 つ値のうち, 4 つは実際に開いたものであったため, ユーザのシチュエーションとして, 閲覧時の場所と時間を利用することは有用ではないかと考えられる. タイトルとユーザのシチュエーションのベクトルを結合した  $\mathbf{X}_t \oplus \mathbf{X}_{st}$ ,  $\mathbf{X}_t \oplus \mathbf{X}_s$  では, タイトル 300 次元に対し, ユーザのシチュエーションが 2 次元または 6 次元と次元数に大きな差があるため, ユーザのシチュエーションの影響が小さくなり,  $\mathbf{X}_t$  とほぼ変わらない予測結果となったと考えられる. ユーザのシチュエーションとして閲覧時の時間ののみを利用した  $f(\mathbf{X}_t) \oplus \mathbf{X}_{st}$  では, 正解データのうち,  $\mathbf{X}_t$  ではピークが目立たなかった箇所 (Index=11,

6 : <https://scikit-learn.org/stable/>

7 : <https://flutter.dev/>

8 : <https://firebase.google.com/>

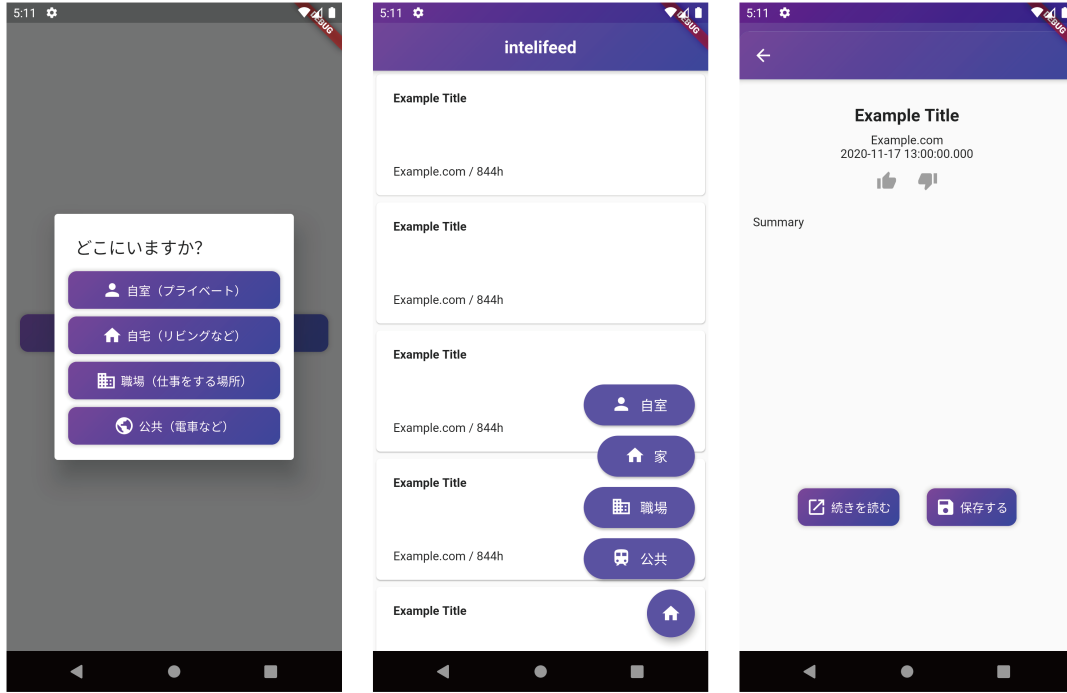


図 2 プロトタイプシステム（左：場所選択画面，中央：ホーム画面，右：記事表示画面）

表 2 各手法とパラメータ

説明変数	説明	Kernel	Degree	$C$	$\epsilon$	$R^2$
1. $\mathbf{X}_t$	タイトルのみ	poly	3	1.0	0.01	0.0261
2. $\mathbf{X}_{st}$	シチュエーション（時間）のみ	poly	5	0.001	0.1	-0.0245
3. $\mathbf{X}_s$	シチュエーション（時間，場所）のみ	poly	5	1.0	0.01	0.1244
4. $\mathbf{X}_t \oplus \mathbf{X}_{st}$	タイトル $\oplus$ シチュエーション（時間）	poly	3	1.0	0.01	0.0268
5. $\mathbf{X}_t \oplus \mathbf{X}_s$	タイトル $\oplus$ シチュエーション（時間，場所）	poly	3	1.0	0.01	0.0279
6. $f(\mathbf{X}_t) \oplus \mathbf{X}_{st}$	1. の結果 $\oplus$ シチュエーション（時間）	poly	1	0.1	0.1	0.5592
7. $f(\mathbf{X}_t) \oplus \mathbf{X}_s$	1. の結果 $\oplus$ シチュエーション（時間，場所）	rbf	-	1.0	0.01	0.5816

71 等）について，予測出来ている箇所があることから，タイトルと閲覧時の時間を組み合わせることによって，ユーザが読むであろう点を予測できる可能性を示唆していると考えられる． $f(\mathbf{X}_t) \oplus \mathbf{X}_s$  は， $\mathbf{X}_s$  と比べ，正解データのうち，ピークが出現する回数が 7 回と多くなったことから，タイトルのベクトルを用いた  $\mathbf{X}_t$  の結果とユーザのシチュエーションを組み合わせることが有用であると考えられる．

今回の評価では，テストデータとして，実際にユーザが開いたコンテンツを正解データとしているため，コンテンツに興味があるにも関わらず，見落とししたりした場合の評価は行えていないため，実際にユーザに予測値を提示して評価する必要がある．

## 5 評価実験

作成したユーザモデルの評価実験を行う．実際に図 2 の実装したアプリケーションを被験者に 5 日間利用してもらい，ログデータを収集する．収集したログデータをもとにクラウドソーシングにて定量的評価を行う．

### 5.1 ログデータ収集

被験者に図 2 のアプリケーションを利用してもらい，各手法での推薦精度の評価のためのデータ収集を行う．対象は 20 代の日常的にインターネットを使用する 8 人を対象として，アプリケーションをインストールしてもらい，5 日間の実験を実施した．

被験者には事前に Feedly に 10 サイト以上のニュースサイトをフォローしてもらい，実験期間中でも，新たにニュースサイトをフォローしても良いと指示した．学習データを収集するため，期間中 1 日 3 回以上アプリケーション上でニュースを確認するよう指示し，リマインドのため 11 時，15 時，23 時にプッシュ通知を送信し，ニュースを確認するよう促した．

推薦手法は，表 2 の  $\mathbf{X}_t$ ， $f(\mathbf{X}_t) \oplus \mathbf{X}_{st}$ ， $f(\mathbf{X}_t) \oplus \mathbf{X}_s$  の 3 つ，ベースラインとして被験者が表示した記事の中から，未読の記事をランダムに提示する Random の合計 4 手法を用いた．各手法のモデル作成は午前 0 時と午後 0 時に行い，推薦記事は，被験者が期間中，記事を 150 件以上表示した場合のみ，予測値が 0.1 以上の上位最大 5 件を提示した．被験者ごとに推薦手法をランダムに割り当て（各手法 2 人），実験期間中，推薦手法は変更しない．

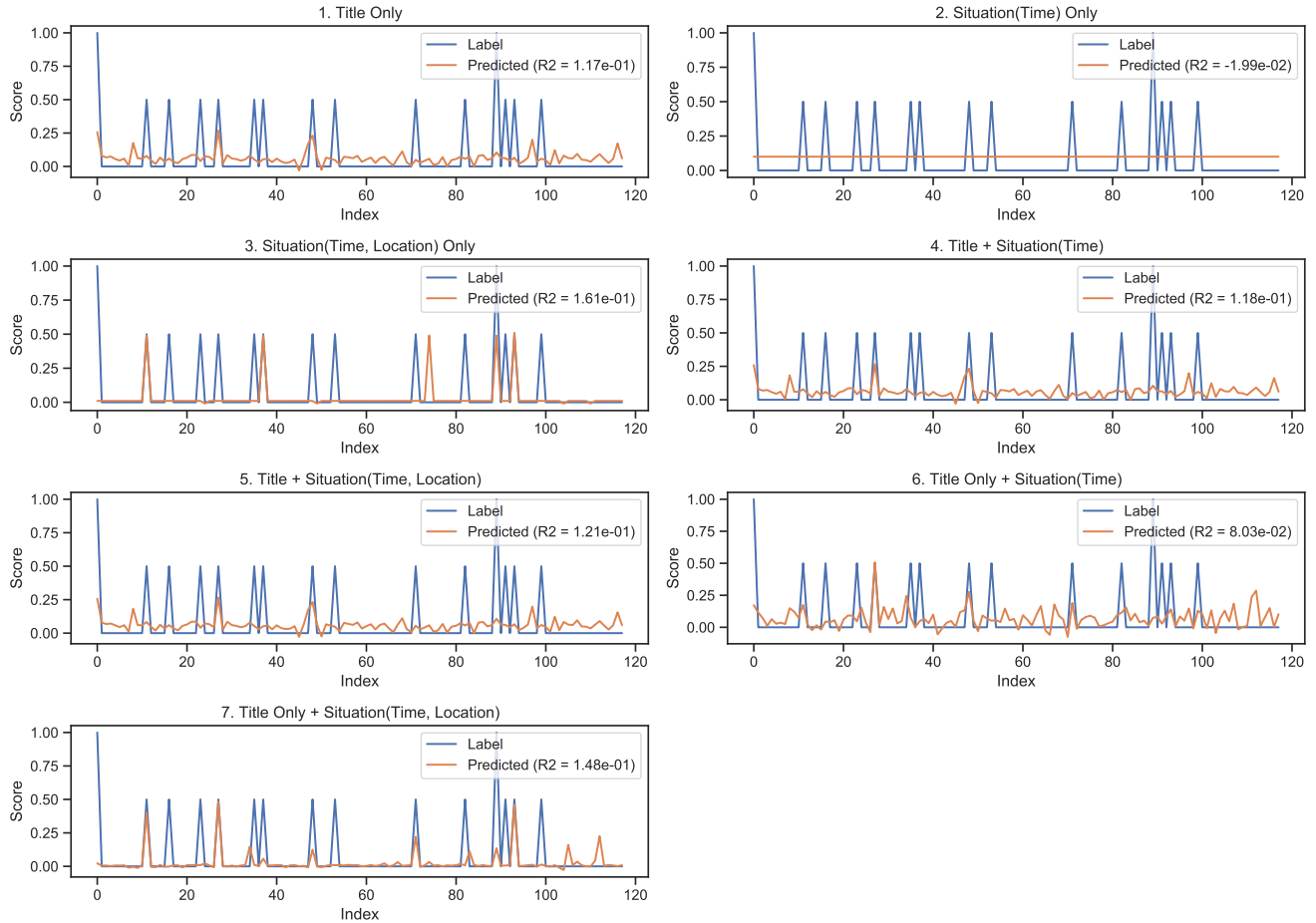


図3 予測結果

収集したデータを表3に示す。今回実験期間が、5日間という短い期間であったため、推薦記事が提示されるまで時間がかり、十分に推薦記事を提示することができなかった。そのため、評価は収集したログデータを元にクラウドソーシングを用いたフォームでの定量的評価にて行う。

## 5.2 定量的評価

収集したログデータをもとに定量的評価を行う。はじめに、被験者に、あるユーザの閲覧ログを確認してもらう。その後、自宅(home)15:00におけるRandom,  $\mathbf{X}_t$ ,  $f(\mathbf{X}_t) \oplus \mathbf{X}_{st}$ ,  $f(\mathbf{X}_t) \oplus \mathbf{X}_s$  各手法の結果上位2件、合計8件提示し、被験者が閲覧ログのユーザであると仮定し、読みたいと思うかどうかを5段階で評価してもらった。フォームの一例を図5に示す。

ユーザごとの閲覧ログは、図4に示すように、要約文を開いた記事の中からランダムで選択した4件提示する。また、ユーザごとに閲覧ログは5つ用意した。なお表3のユーザDについては要約文開封数が4のため1つのみである。被験者は20代の学生11名とクラウドソーシングサービスであるCrowdWorks<sup>9</sup>で募集した。被験者はユーザごとに1つの閲覧ログに対応するフォームに回答してもらう。各フォームごとの有効な回答数を表4に示す。

## 5.3 結果と考察

各手法の平均値と分散を表5、図6に示す。平均値は、 $f(\mathbf{X}_t) \oplus \mathbf{X}_s$ ,  $\mathbf{X}_t$ ,  $f(\mathbf{X}_t) \oplus \mathbf{X}_{st}$ , Randomの順に高いことが確認できた。また、学生とクラウドソーシングの被験者間で、平均値の大小関係の違いはなかった。(表6)

続いて、分散分析を行った。有意水準0.05において、One-way ANOVAの結果、 $p = 6.722 \times 10^{-7}$ となり、有意差が確認できた。また、Tukey HSDの結果(表7)、Randomと $\mathbf{X}_t$ 間では、 $p = 0.024$ と有意差が確認できた。Randomと $f(\mathbf{X}_t) \oplus \mathbf{X}_{st}$ 間では、 $p = 0.868$ と有意差が確認できなかった。Randomと $f(\mathbf{X}_t) \oplus \mathbf{X}_s$ 間では、 $p = 0.001$ と有意差が確認できた。 $f(\mathbf{X}_t) \oplus \mathbf{X}_{st}$ と $f(\mathbf{X}_t) \oplus \mathbf{X}_s$ 間では、 $p = 0.001$ と有意差が確認できた。 $\mathbf{X}_t$ と $f(\mathbf{X}_t) \oplus \mathbf{X}_s$ 間では、 $p = 0.102$ と有意差が確認できなかった。Randomと各手法の平均値の差が小さいのは、Feedlyの特性上、ユーザが好みのニュースサイトからのみ記事を表示することが原因であると考えられる。また、Randomと $\mathbf{X}_t$ 間で有意差が確認できて、Randomと $f(\mathbf{X}_t) \oplus \mathbf{X}_{st}$ 間では有意差が確認できなかったことから、SVRの説明変数として、ユーザが記事を読んだ日時を用いることは不適切であると考えられる。 $f(\mathbf{X}_t) \oplus \mathbf{X}_s$ において、 $\mathbf{X}_t$ との有意差は確認できなかったが、平均値が上回っているため、SVRの説明変数として、ユーザが記事を読んだ場所を用い

9: <https://crowdworks.jp/>

表 3 収集データ

ユーザ	手法	合計記事取得数	要約文開封数	全文開封数	記事推薦数	推薦記事開封数
A	$f(\mathbf{X}_t) \oplus \mathbf{X}_{st}$	503	34	27	17	3
B	$\mathbf{X}_t$	834	9	1	0	0
C	$f(\mathbf{X}_t) \oplus \mathbf{X}_s$	974	24	16	12	0
D	Random	701	4	3	29	1
E	$\mathbf{X}_t$	419	32	22	11	0
F	Random	326	21	20	0	0
G	$f(\mathbf{X}_t) \oplus \mathbf{X}_{st}$	204	14	7	0	0
H	$f(\mathbf{X}_t) \oplus \mathbf{X}_s$	807	46	28	22	0

## 認証コード：PJYqb

時刻	場所	サイト	タイトル	URL
15:10:42.648000	home	NHKニュース	"お粗末"避難訓練に市長が立腹 改めて抜き打ち訓練へ	<a href="http://www3.nhk.or.jp/news/html/20210115/k10012817171000.html">http://www3.nhk.or.jp/news/html/20210115/k10012817171000.html</a>
15:10:38.494000	home	朝日新聞デジタル	ネット代などの社員補助、半額非課税に テレワーク促す	<a href="http://www.asahi.com/articles/ASP1H71YLP1HULFA00S.html?ref=rss">http://www.asahi.com/articles/ASP1H71YLP1HULFA00S.html?ref=rss</a>
15:09:12.334000	home	朝日新聞デジタル	半数近くが部屋着のまま出席 オンライン会議で千人調査	<a href="http://www.asahi.com/articles/ASP1H7JZGP16UZPS006.html?ref=rss">http://www.asahi.com/articles/ASP1H7JZGP16UZPS006.html?ref=rss</a>
01:27:34.516000	home	jiji.comアクセスランキング	初確認のコロナ変異種、日本に情報提供要請 ブラジル政府	<a href="https://www.jiji.com/jc/article?k=2021011100164&amp;g=int&amp;m=rss">https://www.jiji.com/jc/article?k=2021011100164&amp;g=int&amp;m=rss</a>

図 4 閲覧ログページ

ることで推薦精度が向上する可能性があると考えられる。

各手法の平均値とユーザの記事表示数を図 7 に示す。記事表示数と推薦精度の関係性を得ることはできなかった。しかし、 $f(\mathbf{X}_t) \oplus \mathbf{X}_s$  において、記事表示件数が 200 件程度でも十分な推薦結果が得られることがわかった。また記事表示数が 807 のユーザ H の評価値が特に低いことが確認できる。これは、推薦記事や閲覧ログに一般的でない専門的な記事が含まれており、被験者が適切に判断することができなかったことが原因であると考えられる。特に  $f(\mathbf{X}_t) \oplus \mathbf{X}_{st}$  において平均値が 1.700 と低く、推薦内容は“最近の質問 - スタック・オーバーフロー、Realtime Database transaction disconnect エラー”<sup>10</sup> と“最近の質問 - スタック・オーバーフロー、anaconda navigator が起動しない。再インストールも効果なし。”<sup>11</sup> であった。

実際のユーザの閲覧ログを用いた定量的評価の結果、SVR を用いたユーザモデル作成においてユーザのシチュエーションとして、ユーザのコンテンツを閲覧した場所、コンテンツの内容としてニュース記事のタイトルを目的変数として利用することが効果的であることが示唆された。

## 6 おわりに

Web クリップアプリ、ニュースフィードアプリ等でユーザが見落とししたり、保存したこと自体を忘れたコンテンツをユーザに適切なタイミングで、推薦することによって、ユーザの知識

獲得を支援することは可能であるかを研究課題とし、本研究では、コンテンツ推薦に必要なユーザモデルの作成を行った。ユーザモデルは、ニュースタイトルの分散表現を SVR で学習した結果と、ユーザのシチュエーション（時間、場所）を結合したベクトルを説明変数、ユーザの閲覧行動を目的変数として、SVR で学習し作成し、評価を行った。また、作成したユーザモデルを用いて、ユーザの現在のシチュエーションで、コンテンツを閲覧するには適切かを判定し、適切である場合に、過去に取得したユーザの興味あると予測された未閲覧のコンテンツを提示した。実際のユーザの閲覧ログを用いた定量的評価の結果、SVR を用いたユーザモデル作成において、ユーザのシチュエーションとして、ユーザのコンテンツを閲覧した場所、コンテンツの内容としてニュース記事のタイトルを目的変数として利用することが効果的であることが示唆された。今後の課題として、提案システムを実装したシステムの長期的な評価実験による評価と、ユーザの閲覧場所の自動推定が挙げられる。さらに、ユーザの記事の閲覧順序や閲覧時間などのコンテキスト考慮した手法や、アプリケーションの起動タイミングと起動時間などを元に、ユーザが記事を閲覧することができるような提示方法について検討する。また、コンテンツとして今回ニュース記事を用いたが、動画コンテンツなどの閲覧する際にシチュエーションによる制限が多いコンテンツなどに適用する手法について検討する。

10 : <https://ja.stackoverflow.com/questions/72773>

11 : <https://ja.stackoverflow.com/questions/72779>

## ニュース記事に関する調査

**\*必須**

クラウドワークスIDを入力してください \*

回答を入力

あるユーザーが読んだニュース記事の履歴です。URLを確認してニュース記事を読んでください。

[https://kitayama-lab.github.io/intelifeed-history/4mjeNUypF6XQtmjcw0EKZ2oAu32\\_1](https://kitayama-lab.github.io/intelifeed-history/4mjeNUypF6XQtmjcw0EKZ2oAu32_1)

場所について

- private: 自宅（プライベートな場所）
- home: 自宅（リビングなど）
- work: 職場
- public: 公共の場所（電車、公園など）

あなたは履歴中のニュース記事に興味ありますか？ \*

1 2 3 4 5

まったく興味がない ☐ ☐ ☐ ☐ ☐ とても興味がある

この履歴のユーザーになりきってください。自宅 15:00 というシチュエーションで、以下の記事が、提示された場合、読みたいと思いますか？

AFPBB News - 総合新着記事100 -, マスクより信仰が身を守る ガンジス川で多数が沐浴 インド \*

<https://www.afpbb.com/articles/-/3326508>

1 2 3 4 5

読みたいと思わない ☐ ☐ ☐ ☐ ☐ とても読みたいと思う

Newsweek ニュース速報, 6%の21年仏成長率予想は達成可能、後半に伸び加速へ＝財務相 \*

<https://www.newsweekjapan.jp/headlines/business/2021/01/308699.php>

1 2 3 4 5

読みたいと思わない ☐ ☐ ☐ ☐ ☐ とても読みたいと思う

図 5 質問内容一部

## 謝 辞

本研究の一部は、2020 年度科研費基盤研究 (C)(課題番号：18K11551) によるものです。ここに記して謝意を表すものとします。

## 文 献

- [1] 土岐真里奈, 牛尼剛聡. ソーシャルストリーム閲覧時の振舞いを利用したユーザプロファイル構成手法. 情報処理学会論文誌データベース (TOD) , Vol. 6, No. 4, pp. 35–45, Sep 2013.
- [2] Rui Liu, Huilin Peng, Yong Chen, and Dell Zhang. Hypernews: Simultaneous news recommendation and active-time prediction via a double-task deep neural network. pp. 3459–3465, Jul 2020.
- [3] Jon Gulla, Lemei Zhang, Peng Liu, Özlem Özgöbek, and Xiaomeng Su. The adressa dataset for news recommendation. pp. 1042–1048, Aug 2017.
- [4] 小林鼓, 藤田桂英. ユーザの過去ツイートをを用いた噂の早期検

表 4 フォームごとの有効回答数

ユーザ	フォーム ID	有効回答数	合計
A	1	36	
A	2	20	
A	3	20	
A	4	19	
A	5	18	113
B	1	30	
B	2	20	
B	3	20	
B	4	22	
B	5	20	112
C	1	30	
C	2	20	
C	3	18	
C	4	21	
C	5	20	109
D	1	29	29
E	1	27	
E	2	19	
E	3	20	
E	4	18	
E	5	21	105
F	1	33	
F	2	20	
F	3	17	
F	4	19	
F	5	24	113
G	1	26	
G	2	20	
G	3	19	
G	4	20	
G	5	18	103
H	1	27	
H	2	19	
H	3	20	
H	4	20	
H	5	16	102

表 5 各手法の平均値と分散

手法	平均	分散
Random	2.831	0.218
$\mathbf{X}_t$	2.949	0.442
$f(\mathbf{X}_t) \oplus \mathbf{X}_{st}$	2.910	0.613
$f(\mathbf{X}_t) \oplus \mathbf{X}_s$	3.092	0.244

出. 研究報告知能システム (ICS) , Vol. 2020-ICS-199, No. 9, Mar 2020.

- [5] Yang Liu and Yi-Fang Wu. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. Dec 2018.
- [6] 奥健太, 中島伸介, 宮崎純, 植村俊亮, 加藤博一. ユーザの時系列コンテキストを考慮した情報推薦方式の提案. 情報処理学会



表 6 被検者ごとの各手法の平均値

	Random	$\mathbf{X}_t$	$f(\mathbf{X}_t) \oplus \mathbf{X}_{st}$	$f(\mathbf{X}_t) \oplus \mathbf{X}_s$
学生	2.909	3.327	3.163	3.459
クラウドソーシング	2.818	2.982	2.840	3.071

表 7 Tukey HSD 結果

group 1	group 2	p
Random	$\mathbf{X}_t$	0.024
Random	$f(\mathbf{X}_t) \oplus \mathbf{X}_{st}$	0.868
Random	$f(\mathbf{X}_t) \oplus \mathbf{X}_t$	0.001
$\mathbf{X}_t$	$f(\mathbf{X}_t) \oplus \mathbf{X}_t$	0.102
$f(\mathbf{X}_t) \oplus \mathbf{X}_{st}$	$f(\mathbf{X}_t) \oplus \mathbf{X}_t$	0.001

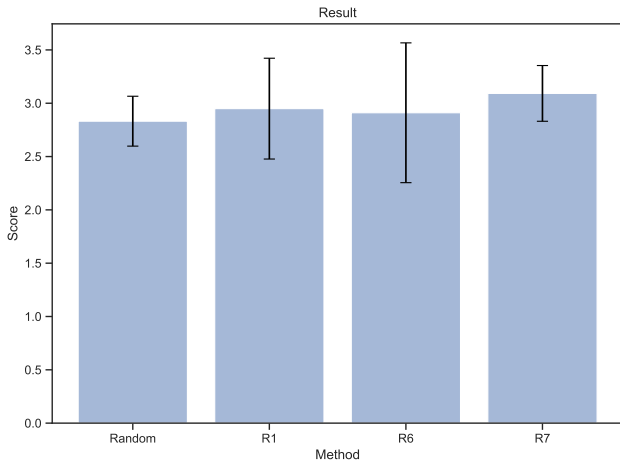


図 6 各手法の結果

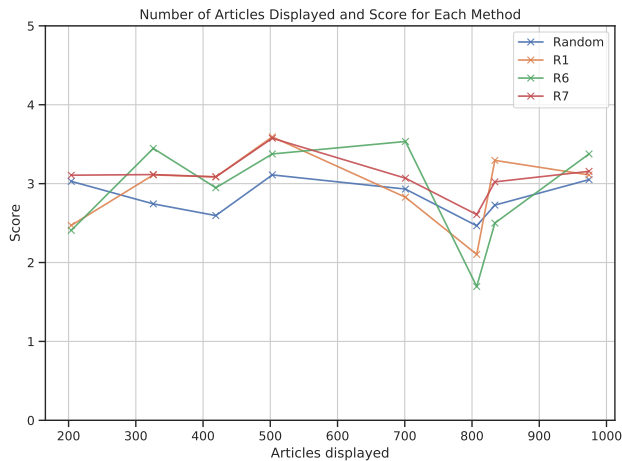


図 7 各手法の平均値と記事表示数

研究報告. データベース・システム研究会報告, Vol. 146, pp. 121–126, Sep 2008.

- [7] 中村明順, 西尾信彦. 超個人化プロフィール生成のための web ライフログの分類分析. 電子情報通信学会技術研究報告. MoMuC, モバイルマルチメディア通信, Vol. 111, No. 476, pp. 135–142, Mar 2012.
- [8] Marios Constantinides and John Dowell. A framework for interaction-driven user modeling of mobile news reading behaviour. In *Proceedings of the 26th Conference on User*

*Modeling, Adaptation and Personalization*, UMAP '18, p. 33–41, New York, NY, USA, 2018. Association for Computing Machinery.

- [9] Marios Constantinides, Panagiotis Germanakos, George Samaras, and John Dowell. Your digital news reading habits reflect your personality. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*, UMAP '18, p. 45–48, New York, NY, USA, 2018. Association for Computing Machinery.
- [10] Azamat Bolat, Dongmin Kim, and Ki-Joune Li. Discovering user-context in indoor space. In *Proceedings of the 9th ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness*, ISA'18, p. 1–6, New York, NY, USA, 2018. Association for Computing Machinery.
- [11] Shuochao Yao, Shaohan Hu, Yiran Zhao, Aston Zhang, and Tarek F. Abdelzaher. Deepsense: A unified deep learning framework for time-series mobile sensing data processing. *CoRR*, Vol. abs/1611.01942, , 2016.
- [12] Y. Watanabe, R. Suzumura, S. Matsuno, and M. Ohyama. Investigation of context-aware system using activity recognition. In *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*, pp. 287–291, Feb 2019.
- [13] Daniel Ashbrook and Thad Starner. Starner, t.: Using gps to learn significant locations and predict movement across multiple users. *personal and ubiquitous computing* 7(5), 275–286. *Personal and Ubiquitous Computing*, Vol. 7, pp. 275–286, Oct 2003.
- [14] 笠井昭範, 原直, 阿部匡伸. Flag : 位置情報を基軸としたライフログ集約システム (サービスコンピューティング). 電子情報通信学会技術研究報告, Vol. 114, No. 157, pp. 29–34, Jul 2014.
- [15] T. KUDO. Mecab : Yet another part-of-speech and morphological analyzer. <http://mecab.sourceforge.net/>, 2005.
- [16] Tomas Mikolov, Edouard Grave, Piotr Bojanowski, Christian Puhresch, and Armand Joulin. Advances in pre-training distributed word representations. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.