情報処理学会論文誌　データベース　Vol.15　No.1　1–11 (Jan. 2022)

実空間のユーザ行動分析に基づく潜在的興味分析方式

大村 貴信[1,a)]　鈴木 健太[1,b)]　パノット シリアーラヤ[2,c)]　栗 達[3,d)]
河合 由起子[3,e)]　中島 伸介[3,f)]

受付日 2021年6月8日, 採録日 2021年10月4日

概要：携帯端末向けの Web 広告サービスは年々増加傾向にあり，検索キーワードや閲覧履歴，SNS への投稿内容によるユーザの操作履歴に基づく分析手法の研究開発が広く行われている．また携帯端末の GPS から得られるユーザの位置情報を利用し，ユーザの現在地や居住地区に合わせた広告を推薦する手法の開発も行われている．しかしながら，これら位置情報の軌跡から得られる実空間の行動と Web 閲覧や SNS 投稿の操作履歴の両データを用いた分析手法は，実空間の位置・時刻における操作履歴より抽出された特徴語との相関に基づいた明示的興味分析にとどまっており，周辺環境の影響は十分に考慮されず，ユーザの潜在的興味の抽出には至っていない．そこで本研究では，実空間における時空間での行動範囲と興味範囲を拡張することで，ユーザの潜在的興味を行動履歴とジオタグ付き SNS データから推定する手法を提案する．本研究では，携帯端末での Web 広告推薦への応用を目指し，ユーザの興味対象を実空間に存在する店舗とする．具体的には，まず，ユーザ行動ログデータとジオタグ付きツイートデータからユーザが訪問した店舗の時間と場所に基づき，時間と位置の行動範囲を拡張する．次に，行動範囲内の店舗属性を拡張し，それら拡張された行動範囲と興味範囲に基づき特徴を抽出し学習モデルを生成する．本稿では，拡張した時空間の行動範囲ごとに学習モデルを生成し，特定の店舗を訪問するユーザの予測精度を検証する．

キーワード：潜在的興味分析，推薦システム，ユーザ行動分析，広告推薦

Latent Interest Analysis Utilizing User Movement History in Real-space

Takanobu Omura[1,a)]　Kenta Suzuki[1,b)]　Panote Siriaraya[2,c)]　Da Li[3,d)]　Yukiko Kawai[3,e)]
Shinsuke Nakajima[3,f)]

Received: June 8, 2021, Accepted: October 4, 2021

Abstract: Web advertising services for mobile devices are rapidly increasing. Generally, the methods of advertisement recommendation are according to the analysis of user behavior such as searching keywords and keyword matching based on browsing history. On the other hand, a great number of researches also focused on the location information which are obtained from GPS of the mobile devices. These advertisement recommendation methods are developed based on the users' current location and residential area. However, both of the above two approaches did not take into account the impact of the surrounding environment, consequently, it is difficult to extract the potential interests of users. Therefore, in this paper, we firstly extend the range of movement and interest in the spatio-temporal real-space. Secondly, we propose a latent interest detection method based on user movement history and the geotagged social media data. Especially, because we focus on the application of Web advertisement recommendation for mobile devices, the aim of user interest is defined as commercial facility in this research. Finally, we increase the commercial facility attributes number utilizing the extended user movement range, and apply a neural network approach for extracting user interest features. In addition, to verify the validity of our proposed method, we evaluate the neural network models and verify the accuracy of predicting if the users will visit a particular store.

Keywords: latent interest analysis, recommendation system, user movement history, advertisement recommendation

1. はじめに

インターネット向けの Web 広告サービスは年々増加傾向 [1] にあり，検索キーワードや閲覧履歴，SNS への投稿内容によるユーザの操作履歴に基づく Web 広告推薦手法の研究開発が広く行われている．また携帯端末の GPS から得られるユーザの位置情報を利用し，ユーザの現在地や居住地区に合わせた広告を推薦する手法の開発も行われている．しかしながら，これら位置情報の軌跡から得られる実空間の行動と Web 閲覧や SNS 投稿の操作履歴の両データを用いた分析手法は，実空間の位置・時刻における操作履歴より抽出された特徴語との相関に基づいた明示的興味分析にとどまっており，周辺環境の影響は十分に考慮されておらず，ユーザの潜在的興味の抽出には至っていない．そこで本研究では，実空間における時空間での行動範囲と興味範囲を拡張することで，ユーザの潜在的興味を行動履歴とジオタグ付き SNS データから推定する手法を提案する．なお，将来的には実空間にて活動中のユーザに対する実店舗の広告推薦への応用を検討しており，ある広告対象店舗への訪問履歴がないユーザの行動分析を行うことで，この広告対象店舗への訪問予測，すなわち対象店舗への潜在的興味分析が可能な手法の開発を目的としている．

具体的な手法としては，ユーザ行動ログデータとジオタグ付きツイートデータから，ユーザが訪問した店舗の時間と場所に基づき，時間と位置の行動範囲の概念を拡張する．次に，行動範囲内の店舗属性を拡張し，それら拡張された行動範囲と興味範囲の特徴を抽出し学習モデルを生成する．本稿では，拡張した時空間の行動範囲ごとに学習モデルを生成し，特定の店舗を訪問するユーザの予測精度を検証する．

本研究は一般にジオターゲティング [2], [3] といわれる手法の 1 つに位置付けられると考える．ジオターゲティングとはユーザの位置情報を利用したマーケティング手法である．ユーザの現在地や居住地区に合わせた広告推薦が可能であり，Web 広告を通じて実空間に存在する店舗への来店につなげるという魅力がある．ただし，従来のジオターゲティングの多くは，基本的に実空間の位置情報を利用するものであるのに対して，提案手法ではユーザの実空間での行動に対する意味的な分析をあわせて行うものであり，独自性・新規性は高いと考えている．

本稿の構成は以下のとおりである．2 章では関連研究を紹介する．3 章では提案手法について詳細を説明する．4 章では実験の条件，結果，考察について述べる．最後に 5 章でまとめを記述する．

2. 関連研究

2.1 広告の推薦に関する研究

広告の CV（コンバージョン）率を上げるために様々な研究がなされている．ユーザが次に見たい情報を予測し，それに関する広告を配信するシステムを開発・検証した研究 [4] や消費者が必要とする商品情報とデザインおよびメッセージを個人に合わせたインターネット広告の構成手法を提案している研究 [5]，閲覧行動パターンを考慮した購買予兆の発見モデルを提案している研究がある．また長期的興味と短期的興味を考慮したユーザの潜在的興味分析手法の提案・検証を行った研究 [6] や長期的な経験と直近の経験を考慮するため，生涯シーケンシャルモデリングを用いた研究 [7] がなされている．これらの研究は閲覧履歴などのユーザの Web 空間の情報を用いることで CV 率を上げる研究を行っている．本研究ではユーザの実空間での行動履歴を用いて潜在的興味を推定し広告を推薦することで，クリックや EC サイトでの購入といった Web 空間での CV に加え，実店舗への来店という CV も上げることができると考えている．

2.2 POI 予測に関する研究

ユーザが次に訪れる Point of Interest（POI）を予測および推薦する研究もさかんに行われている．ユーザの過去の行動履歴から次の行動を予測し，POI 推薦するためのジオトピックモデルを提案している研究がある [8]．これは食べログ[*1]の店舗への訪問履歴（レビュー履歴）と Flickr[*2]の写真のジオタグ情報を用いて行動履歴を再現し，人間が行動するときの特徴を用いた POI 推薦を行っている．また従来の POI 推薦の欠点である「ユーザベース協調フィルタリングではユーザの好みが十分に考慮されない」，「地理的な影響力をモデル化する場合，地理的特徴が深く検討されていない」という 2 つの問題を解決するための新しい POI 推薦アプローチを提案している研究がある [9]．これは Gowalla[*3]のデータを使用し，協調フィルタリングと地理的特徴を組み合わせて POI 推薦を行っている．ほかには Location-Based Social Networks（LBSNs）のチェックイン記録が疎であるため，POI 予測および推薦するこ

1 京都産業大学大学院先端情報学研究科
Division of Frontier Informatics, Kyoto Sangyo University, Kyoto 603–8555, Japan
2 京都工芸繊維大学情報工学・人間科学系
Information and Human Science, Kyoto Institute of Technology, Kyoto 606–8585, Japan
3 京都産業大学情報理工学部
Faculty of Information Science and Engineering, Kyoto Sangyo University, Kyoto 603–8555, Japan
a) i2186023@cc.kyoto-su.ac.jp
b) i2086060@cc.kyoto-su.ac.jp
c) spanote@kit.ac.jp
d) lida@cc.kyoto-su.ac.jp
e) kawai@cc.kyoto-su.ac.jp
f) nakajima@cc.kyoto-su.ac.jp

*1 https://tabelog.com/
*2 https://www.flickr.com/
*3 https://go.gowalla.com/

とが難しいという問題を解決するために，ユーザチェックイン行動のシーケンシャルパターンをキャプチャするモデル，VANext（Variation Attention based Next）を提案している研究 [10] や文脈的特徴（時間帯，曜日，場所のカテゴリなど）から学習した，パーソナライズされた潜在行動パターンを活用し，推薦の効果を向上させる 2 種類のモデルを提案している研究 [11] もある．これらの研究は Foursquare[*4]や Gowalla のデータを用いて POI 推薦を行っている．本研究では関連研究でも扱っている LBSNs データ（ジオタグ付きツイートデータ）とユーザ行動ログデータを用いて検証・評価を行うことで，明示的に投稿したかそうでないかの違いが，興味推定にどの程度影響するのか確かめることができると考える．また関連研究では人間が行動するときの特徴の使用や POI 推薦の問題点を解決できるモデルの作成で POI 予測の精度を向上させているが，本研究では移動軌跡の周辺情報も利用することで，POI 予測を精度を向上させることができ，これまで広告を推薦できなかったようなユーザにも効果的な広告推薦を行える可能性があると考えている．

3. 実空間のユーザ行動分析に基づく潜在的興味分析方式

本章では，実空間のユーザ行動分析に基づく潜在的興味分析方式の概要を解説し，特徴抽出方法，学習方法，評価方法について説明する．

図 1 に提案システムの概要を示す．本研究では，各ユーザが予測対象となる店舗（予測対象店舗）に訪れるか否かを学習し，分類器を作成する．分類器に未知のユーザ X の行動ログを入力したとき，ユーザ X が予測対象店舗に訪れるか否か判定する．ユーザ X が予測対象店舗に訪れたユーザ群と類似したエリア内を行動しており，予測対象店舗に訪れると分類器に判定された場合，予測対象店舗の広告推薦を行うシステムの開発を将来的な目標としている．**図 2** に本研究にて将来的に開発を目指している推薦システムの推薦例を示す．ユーザが，日常的に「カフェ」や「猫がいるペットショップ」を訪問しているような場合，これら実空間での行動分析に基づいて，このユーザは潜在的には「猫カフェ」にも興味を持つであろうと推定し，近くの「猫カフェ」の広告を推薦することなどが可能になると考えている．

従来の広告推薦では，頻繁に利用する店舗やアイテムを推薦したり，性別や年齢に応じて該当しそうな店舗やアイテムを推薦したりといった比較的単純な手法が採用されているが，広告主が購買層を広げるという意味ではその効果が十分とはいえない．一方，提案手法では行動した周辺エリアの店舗カテゴリを考慮した潜在的な興味分析を行う．

[*4] https://foursquare.com/

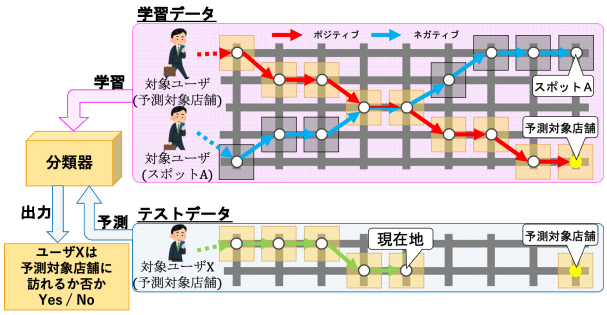

図 1　提案手法のシステム概要
Fig. 1　Overview of the proposed system.

図 2　開発を目指している推薦システムの推薦例
Fig. 2　An example of the final goal of our recommendation system.

これにより，これまで広告を推薦できなかったようなユーザにも効果的な広告推薦を行える可能性があると考えている．

3.1 データ収集およびポジティブ・ネガティブ分類

本研究では 2 種類のデータを使用する．1 つは予測対象店舗の広告配信対象者のユーザ行動ログデータおよび予測対象店舗の広告表示ログデータであり，匿名化されたデータを採用している．なお，本データはユーザ行動ログデータの収集を行う企業よりご提供いただいた．ユーザ行動ログデータは，2019 年のある 1 カ月間に収集された，約 2 億件のデータであり，広告表示ログデータは，同様に 2019 年のある 1 カ月間に収集された約 20 万件のデータである．該当する予測対象店舗は，幅広い年齢層が訪れる日用品・食品なども扱う小売業態の 1 店舗である．選定理由としては，提案手法の分析結果や実験結果が限定的になることがないよう，幅広いユーザが訪れるような店舗を選定した．もう 1 つはジオタグ付きツイートデータであり，2016 年 6 月から 2020 年 6 月までの 4 年間に収集された約 630 万件のデータを採用している．予測対象店舗はユーザ行動ログデータの予測対象店舗と同業態のチェーン 147 店舗を使用している．1 店舗に限定するとデータ量が少なく，分析が難しいと考えたため複数店舗を使用した．**表 1** に基本統計情報を示す．数値はすべてユーザ行動ログデータは 1 カ月間，ジオタグ付きツイートデータは 4 年間のものである．明示的に投稿するジオタグ付きツイートデータとそうでな

表 1　本研究で採用したデータの基本統計情報
Table 1　Basic statistics of the data used in this study.

		ユーザ行動ログデータ	ツイートデータ
期間		1 カ月	4 年間
ユーザ数		115,434	3,414
レコード数		205,061,773	6,269,171
1 人あたりの記録回数	平均	1,776	1,836
	最大	226,779	44,091
	中央	565	1,018
	最小	1	1

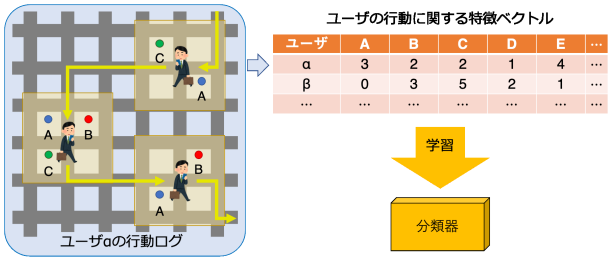

図 3　ユーザ行動特徴ベクトルの抽出方法
Fig. 3　The vector extraction method of the user movement feature.

いユーザ行動ログデータを用いて検証・評価を行うことで，明示的に投稿したかそうでないかの違いが興味推定に，どの程度影響するのか確かめることができると考える．またジオタグ付きツイートデータという比較的入手が容易な疎であるデータとユーザ行動ログデータという入手が困難な密であるデータを比較することで，データの違いがユーザの潜在的興味推定の精度に，どの程度影響があるのかを確認することができると考えるため，これら 2 種類のデータを採用する．

これらのデータから学習用のポジティブデータの候補とネガティブデータの候補となるデータを抽出する．

ジオタグ付きツイートデータでは，あるユーザが予測対象店舗に訪れている場合，そのユーザをポジティブユーザと認定し，このユーザが予測対象店舗に訪れるまでの一定期間のデータを取得し，ポジティブデータの候補とする．またユーザが予測対象店舗に訪れていない場合は，ランダムに選択したツイートを呟くまでの一定期間のデータを取得し，ネガティブデータの候補とした．このときユーザが予測対象店舗に訪れたか否かは，ツイートに含まれている “I’m at starbucks” といった内容から判断した．本研究で使用したジオタグ付きツイートデータはすべて “I’m at ○○”，あるいは “@○○” という表現が含まれているためこの内容を予測対象店舗に訪れたか否かの判断に用いた．

また，ユーザ行動ログデータにはデータに予測対象店舗に滞在しているか否かの項目，広告表示ログデータには広告をクリックしたか否かの項目がある．そのため，少なくとも 1 回以上予測対象店舗に滞在しているユーザデータをポジティブデータの候補とし，逆に広告をクリックしておらず，予測対象店舗に滞在していないユーザデータをネガティブデータの候補とする．

3.2　ユーザ行動特徴ベクトルの抽出

本節では，ユーザ行動特徴ベクトルの抽出について説明する．実空間でのユーザ行動特徴ベクトルの抽出において，本研究では OpenStreetMap [*5] を利用する．OpenStreetMap は，誰でも自由に編集・利用できるオープンな地理情報データである．OpenStreetMap は，無償でカバー率が高く，近年研究に用いられる傾向にあるため本研究でもそのカテゴリ情報を用いた．

図 3 に，ユーザ行動特徴ベクトルの抽出手法を示す．本研究では，ジオタグ付きツイートの投稿場所やユーザ行動ログデータの位置情報検出地点である各記録地点をユーザごとにとりまとめ，この投稿場所や記録地点の周辺スポット情報を OpenStreetMap から取得する．次にこれら周辺スポットをそのカテゴリ（cafe，restaurant，college など）ごとにカウントし，このカウントした情報を基に特徴ベクトルの作成を行う．作成した特徴ベクトルは，ユーザを識別できる ID とカテゴリのカウント情報で構成されている．すなわち，ツイートデータやユーザ行動ログデータに含まれる記録地点の特徴を，周辺に存在するスポットの数やそのバランスによって表現している．このように記録地点の特徴を周辺に存在するスポットのカテゴリおよびその数で表現することで，ユーザがどのような特徴のエリアを訪れたのかを推定することができる．ユーザが訪れた地点から周辺エリアに行動範囲を拡張し，行動範囲内のスポットをカテゴリへと興味範囲を拡張することで，ユーザ行動履歴の特徴表現が可能になると考えている．なお，時系列処理アルゴリズム用の特徴ベクトルでは，さらに単位時間ごとに区切ってカウントしている．

3.3　予測対象店舗への潜在的興味分析手法

予測対象店舗への潜在的興味の分析方法としては，3.2 節で説明したユーザ行動特徴ベクトルを用いて，各種クラス分類手法に基づく学習を行い，予測対象店舗を訪れるユーザモデルを分類器として構築する．この分類器に未知のユーザの行動ログ（特徴ベクトル）をテストデータとして与えたとき，予測対象店舗を訪れるか否かを判定することが可能となる．

4.　潜在的興味分析手法に対する評価実験

本研究では，実空間のユーザ行動分析に基づく潜在的興味分析として，過去にある特定店舗への訪問履歴がないユーザが，今後この店舗を訪問するか否かを推定する手法

*5　https://www.openstreetmap.org/

表 2　本評価実験で採用したポジティブ・ネガティブの候補データの統計情報
Table 2　Statistics of bi-polarity candidate data used in our evaluation experiment.

	ユーザ行動ログデータ	ツイートデータ
期間	1 カ月	4 年間
ユーザ数	847	1,485
レコード数	1,246,697	92,729
期間中の 1 人あたり平均記録回数	1,472	62,444
期間中の 1 人あたり平均スポット数	39,348	1,430

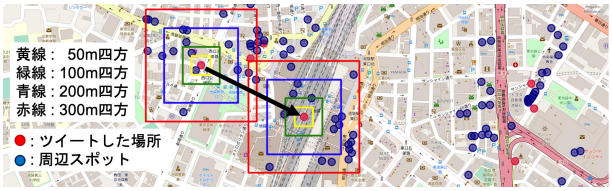

図 4　ユーザ行動特徴抽出のための周辺スポット（青丸）検出
Fig. 4　Peripheral spot (blue points) detection for user behavior feature extraction.

を提案している．本章では，この対象店舗訪問予測手法の精度評価として，各種非時系列処理アルゴリズムによる興味分析と，各種時系列処理アルゴリズムによる興味分析による評価実験を行った．この中で，周辺スポットを抽出するエリアサイズの比較や，分析期間の比較を行い，対象店舗訪問予測を行ううえで，最も適したアルゴリズムや条件を検証したので報告する．なおジオタグ付きツイートデータは同一チェーン店に対し訪問予測を行っているため各店舗の平均を求めている．本評価実験で採用したポジティブ・ネガティブの候補データの統計情報を**表 2** に示す．

4.1　非時系列処理アルゴリズムによる興味分析

4.1.1　エリアサイズの比較に基づく評価

ユーザ行動特徴ベクトルの生成において，ジオタグ付きツイートの投稿場所やユーザ行動ログデータの位置情報検出地点である各記録地点のスポット情報だけはなく，各地点の周辺スポット情報を含めることで，ユーザがどのような特徴のエリアを訪問したかを表現することを目指している．**図 4** にユーザ行動特徴抽出のための周辺スポットの検出例を示す．ここでは，ユーザの移動履歴を表す各記録地点を中心とした正方形エリア内に存在するスポットのカテゴリを周辺スポット情報として抽出している．なお，このエリアが広すぎるとユーザの移動地点とは関係が薄いエリアを含む可能性が高くなり，狭すぎるとユーザ行動範囲の特徴を表現するに十分なデータが確保できない可能性があるため，適切なエリアサイズの検討は重要である．

周辺スポットを考慮するエリアサイズとしては，300 メートル四方，200 メートル四方，100 メートル四方，50 メートル四方の 4 種類のエリアサイズのデータを使用した．ジオタグ付きツイートデータではポジティブユーザ・ネガティブユーザともに 120 人，合計 240 ユーザの特徴ベクトルを使用した．これらのユーザは，1 カ月に 20〜100 件のツイートを投稿している．ユーザ行動ログデータではポジティブユーザ・ネガティブユーザともに 120 人，合計 240 ユーザの特徴ベクトルを使用した．

ジオタグ付きツイートデータ，ユーザ行動ログデータともに 4：1 の比率で学習データとテストデータに分割した．前処理後，Decision Tree（DT）[12]，Logistic Regression（LR）[13]，Nearest Centroid（NC）[14]，Naïve Bayes-Multinomial（NB-M）[15]，Random Forest（RF）[16]，XGBoost（XGB）[17]，k-Nearest Neighbors [18]，Multilayer Perceptron [19]，Naïve Bayes-Bernoulli [20]，Passive Aggressive Classifier [21]，Perceptron [22]，Ridge Regression [23]，Support Vector Machine（linear）[24]，Support Vector Machine（rbf）[24] の 14 種類の非時系列処理の機械学習アルゴリズムを実装し，F 値で評価を行った．14 種類のアルゴリズムの中からジオタグ付きツイートデータにおいて結果の良かったアルゴリズム 2 つとユーザ行動ログデータにおいて結果の良かったアルゴリズム 2 つ，後述の期間の比較において結果の良かったアルゴリズム 2 つの計 6 つのアルゴリズムの結果を図に示す．

図 5 に，非時系列処理によるジオタグ付きツイートデータのエリアサイズごとの F 値比較を，**図 6**（ユーザ行動ログデータ）に，同じく非時系列処理によるユーザ行動ログデータのエリアサイズごとの F 値比較を示す．図 5 のジオタグ付きツイートデータの結果では，XGBoost の 300 メートル四方が最も結果が良い．また，図 6 のユーザ行動ログデータの結果では，Random Forest の 300 メートル四方が最も結果が良い．また，ジオタグ付きツイートデータに比べユーザ行動ログデータの方が多くのアルゴリズムで良い結果を示している．この一因としてデータの密度の違いがあげられる．特徴ベクトルに使用したスポット数がジオタグ付きツイートデータでは 1 万件から 20 万件であるのに対し，ユーザ行動ログデータでは 30 万件から 1,000 万件であることからジオタグ付きツイートデータはユーザ行動ログデータに比べデータ量が少ないといえる．加えてエリアサイズごとのデータ量の差がジオタグ付きツイートデータ，ユーザ行動ログデータともに 300 メートル四方が最も良い結果を示した一因であると考えられる．ユーザ行動ログデータの 300 メートル四方で良い結果を示した Random Forest と XGBoost は，200 メートル四方および 100 メートル四方でも良い結果を示していることから非時系列処理アルゴリズムの中ではアンサンブル学習が適しているのではないかと考える．

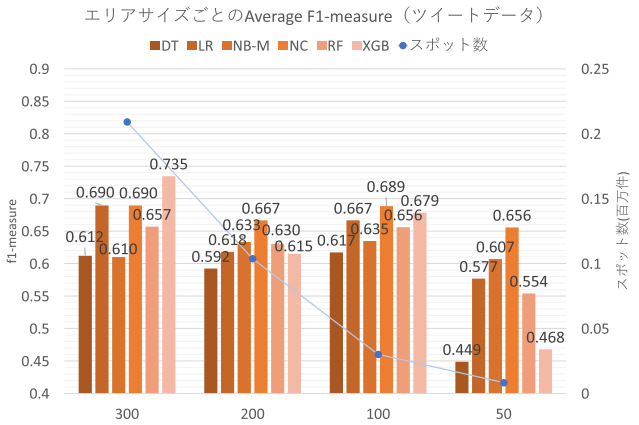

図 5　ジオタグ付きツイートデータのエリアサイズごとの F 値（非時系列処理）

Fig. 5　The results of F-score in each area size on the geotagged tweet data (non-time series processing).

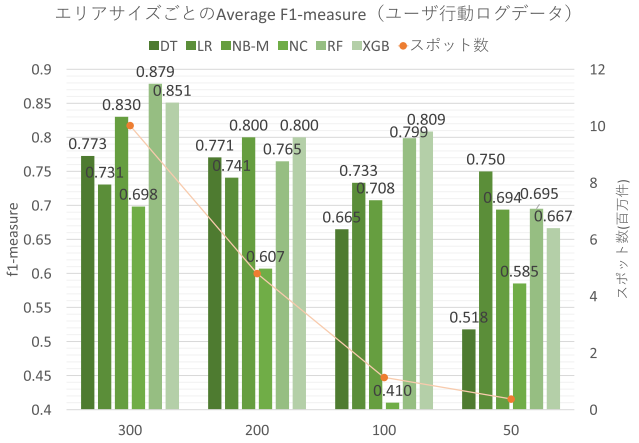

図 6　ユーザ行動ログデータのエリアサイズごとの F 値（非時系列処理）

Fig. 6　The results of F-score in each area size on the user movement log data (non-time series processing).

4.1.2　分析期間の比較に基づく評価

続いて，ジオタグ付きツイートデータの分析における分析期間の比較を行う．分析期間が短すぎると特徴をとらえるために十分な情報が取得できない恐れがあり，長すぎるとユーザの短期的な興味を分析することが困難になると考えられる．なお，ユーザ行動ログデータは，そもそも 1 カ月間のみのデータであるため分析期間の比較は行っていない．分析期間は 1 週間，2 週間，1 カ月，2 カ月，3 カ月の 5 パターンとし，各ユーザのツイート投稿頻度のバラつきを抑えるため，各期間中に**表 3** に示すツイート数を投稿したユーザのみを対象とした．なお，ポジティブユーザ・ネガティブユーザともに 120 人，合計 240 ユーザの特徴ベクトルを使用し，周辺スポットを考慮するエリアサイズは 300 メートル四方とした．

データは 4 : 1 の比率で学習データとテストデータに分割した．前処理後，エリアサイズの比較と同様に非時系列処理の様々な機械学習アルゴリズムを実装し，F 値で評価

表 3　本評価実験で採用したユーザのツイート投稿頻度

Table 3　Frequency of tweets posted.

	各期間中に投稿されたツイート数の範囲
1 週間	5–25
2 週間	10–50
1 カ月	20–100
2 カ月	40–200
3 カ月	60–300

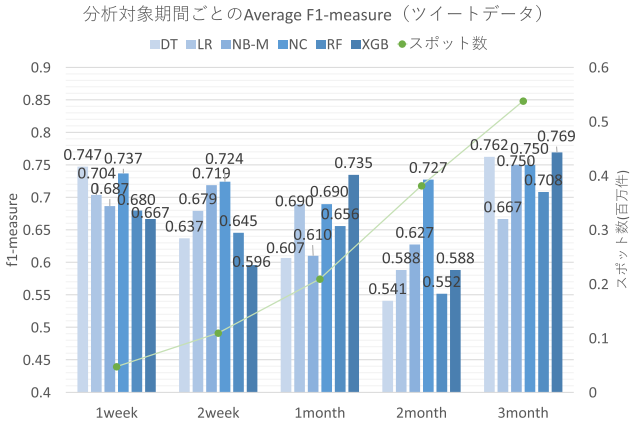

図 7　ジオタグ付きツイートデータの分析対象期間ごとの F 値（非時系列処理）

Fig. 7　The results of F-score in the analysis period on the geotagged tweet data (non-time series processing).

を行った．

図 7 に評価結果を示す．XGBoost の 3 カ月の結果が最も良く，Decision Tree，Nearest Centroid，Naïve Bayes-Multinomial においても 3 カ月が最も良い結果を示している．これは，ある程度長い分析期間を確保することで，ユーザの興味をとらえるにあたって十分なデータ量を確保できたことが要因である可能性がある．しかしながら，特に分析期間が 2 カ月に比べて 1 週間の方が F 値が高くなっている傾向があり，今後他の対象店舗やデータセットに基づく，より詳細な評価実験を行っていく予定である．

4.2　時系列処理アルゴリズムによる興味分析

提案手法ではユーザが予測対象店舗に訪れるか否かを推測することを目指しているが，ユーザの行動はその直前の行動に影響を受けることがある．したがって，ユーザ行動ログの時系列性を考慮することで推測精度を向上させられるかの検証と時系列の順序性の検証を行うため，本節にて時系列処理アルゴリズムによる興味分析の評価を行う．

4.2.1　アルゴリズムの比較に基づく評価

本項では，各種時系列アルゴリズムのうち，提案手法に適用するうえで最も性能が高くなるアルゴリズムを判定するため，アルゴリズムの比較に基づく評価を行う．ここでは，ユーザ行動ログデータを使用し，エリアサイズは 300 メートル四方とした．時系列処理アルゴリズムで使用する

特徴ベクトルの生成方法は，非時系列処理アルゴリズムの特徴ベクトルの生成方法と基本的に同様であるが，異なる点は一定時間の Window サイズごとに区切って，その一定時間内に存在する OSM のカテゴリのカウント情報から生成したベクトルを平均化し，時系列データを作成する点である．ユーザ行動ログデータの全データが 1 カ月分のデータであるためそのすべての期間を考慮した．ここでは，時系列データの区切る Window サイズを 1 日とし，ポジティブユーザ・ネガティブユーザともに 120 人，合計 240 ユーザの特徴ベクトルを使用した．特徴ベクトルを 4 : 1 の比率で学習データとテストデータに分割し，時系列アルゴリズムとしては long short-term memory recurrent neural network（LSTM）[25]，bidirectional LSTM（Bi-LSTM）[26]，attention-based bidirectional LSTM（AttBiLSTM）[27] を用いて，F 値により評価した．これらのアルゴリズムは時系列処理アルゴリズムとしてよく利用されているため本研究でも採用した．評価結果を**図 8** に示す．

図 8 より，時系列処理アルゴリズムの中で AttBiLSTM が最も結果が良い．そこで次節の実験では AttBiLSTM を使用する．

4.2.2　エリアサイズの比較に基づく評価

非時系列処理アルゴリズムと同様に，エリアサイズが広すぎるとユーザの移動地点とは関係が薄いエリアを含む可能性が高くなり，狭すぎるとユーザ行動範囲の特徴を表現するに十分なデータが確保できない可能性があるため，適切なエリアサイズの判定を行うための評価を行う．

エリアサイズとしては，300 メートル四方，200 メートル四方，100 メートル四方，50 メートル四方の 4 種類，かつ時系列データの区切る時間（Window サイズ）を 1 分，10 分，1 時間，10 時間，1 日とした 5 種類の合計 20 種類のデータを使用した．ジオタグ付きツイートデータではポジティブユーザ・ネガティブユーザともに 120 人，合計 240 ユーザの特徴ベクトルを使用した．これらのユーザは，1 カ月に 20〜100 件のツイートを投稿している．ユーザ行動ログデータではポジティブユーザ・ネガティブユーザともに 120 人，合計 240 ユーザの特徴ベクトルを使用した．

ジオタグ付きツイートデータ，ユーザ行動ログデータともに 4 : 1 の比率で学習データとテストデータに分割する．前処理後，attention-based bidirectional LSTM（AttBiLSTM）を用いて，F 値で評価を行った．

図 9 に，ジオタグ付きツイートデータのエリアサイズごとの F 値（時系列処理）を示し，**図 10** に，ユーザ行動ログデータのエリアサイズごとの F 値（時系列処理）を示す．

図 9 では 300 メートル四方の 1 日が最も良く，図 10 では僅差ではあるが 300 メートル四方の 10 時間および 1 日が最も結果が良い．また，ジオタグ付きツイートデータ，ユーザ行動ログデータともにエリアサイズが小さくなるにつれて F 値が下がっていることが確認できる．やはりユーザの興味をとらえるために十分なデータ量を確保できる 300 メートル四方の結果が良くなった可能性があると考える．またジオタグ付きツイートデータにおいては Window サイズは 1 日が最も良い結果となった．ジオタグ付きツイートデータの 300 メートル四方，200 メートル四方，100 メートル四方では Window サイズが大きくなるにつれ結果が良くなっているが，ユーザ行動ログデータでは Widow サイズ間の結果の差がツイートデータほど顕著に見られなかった．これはジオタグ付きツイートデータにポジティブユーザが 120 ユーザしかいなかったため，ランダムに 120 ユーザを選ぶことができなかった．それに対しユーザ行動ログデータはポジティブユーザが約 350 ユーザいたため，その中からランダムに 120 ユーザを選び，学習させ，評価することを 10 回繰り返し，平均を求めることができたからであると考える．そこでユーザ行動ログデータで 10 回繰り返したときの標準偏差を求めた．その結果を**表 4** に示す．表 4 を見ると最大でも 0.04 とばらつきが少ないことが確認できる．120 ユーザランダムに選ぶことを 10 回繰り返したことが Window サイズ間の結果に与えた影響は少ないといえる．このことからジオタグ付きツイートデータではデータ量の確保が重要であったため Window サイズが 1 日の結果が最も良くなり，データ量が十分に確保できているユーザ行動ログデータは Window サイズ間の差が顕著に現れなかったのではないかと考える．今後他の対象店舗に対する評価実験などを継続しながら，引き続き考察していきたいと考えている．

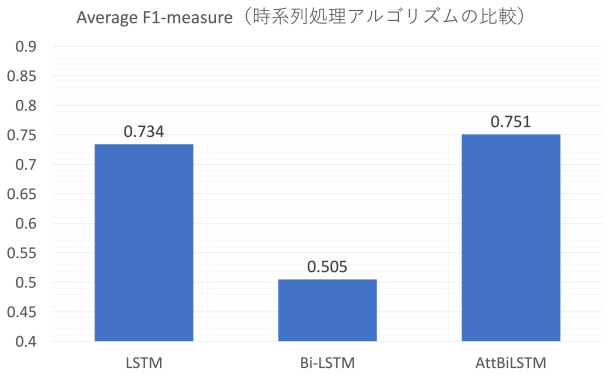

図 8　時系列処理アルゴリズムの比較
Fig. 8　Comparison of time series processing algorithms.

4.3　時系列処理アルゴリズムと非時系列処理アルゴリズムの比較

4.1 節および 4.2 節にて，非時系列処理アルゴリズムおよび時系列アルゴリズムによる興味分析結果について議論した．本節では時系列処理アルゴリズムと非時系列処理アルゴリズムを比較し議論する．

図 11 に，ユーザ行動ログデータのエリアサイズ 300 メートル四方の結果を，時系列処理アルゴリズムおよび非

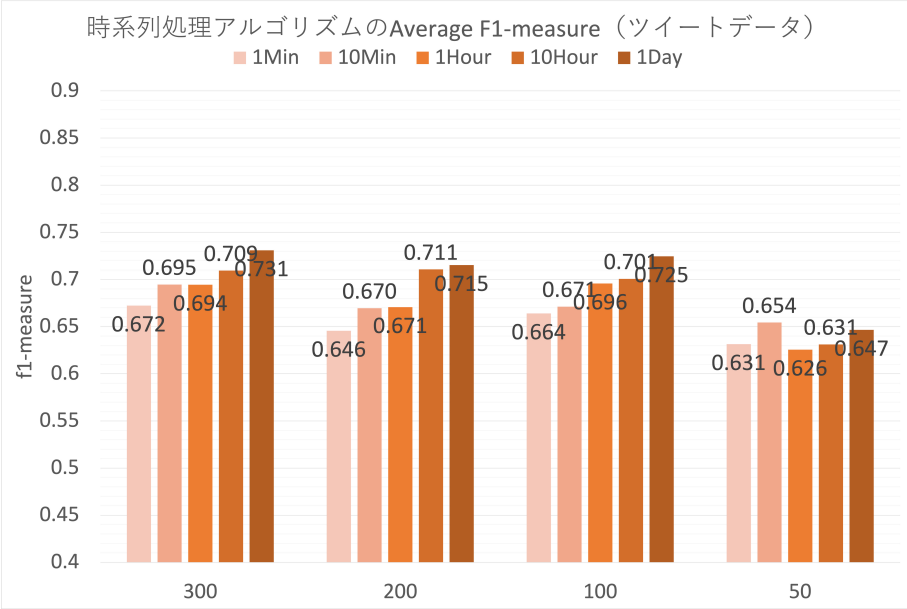

図 9　ジオタグ付きツイートデータのエリアサイズごとの F 値（時系列処理）
Fig. 9　The results of F-score in each area size on the geotagged tweet data (time series processing).

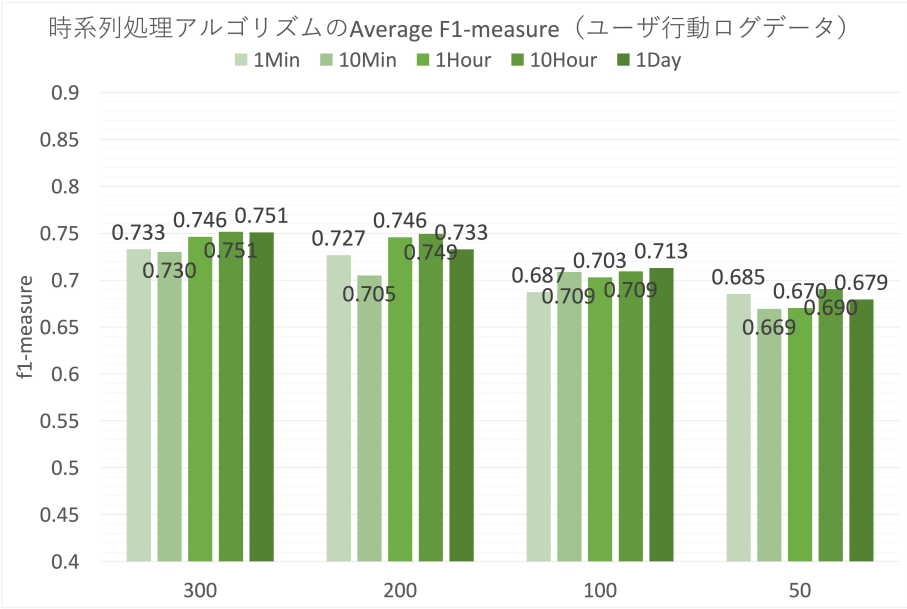

図 10　ユーザ行動ログデータのエリアサイズごとの F 値（時系列処理）
Fig. 10　The results of F-score in each area size on the user movement log data (time series processing).

表 4　ユーザ行動ログデータのエリアサイズ・Window サイズごとの標準偏差（時系列処理）
Table 4　Standard deviations between the each area size on the user movement log data and Window size.

	1 Min	10 Min	1 Hour	10 Hour	1 Day
300	0.03	0.02	0.03	0.03	0.03
200	0.03	0.03	0.03	0.02	0.03
100	0.03	0.04	0.03	0.02	0.04
50	0.02	0.03	0.03	0.01	0.04

時系列処理アルゴリズムで比較したグラフを示す．図 11 が示すとおり，非時系列処理アルゴリズムと時系列処理アルゴリズムを比較した場合，非時系列処理アルゴリズムである Random Forest の F 値の値が良く，XGBoost，Naïve Bayes-Multinomial と続く．このことから今回の実験条件においては非時系列処理アルゴリズムの特にアンサンブル学習が最も良い結果を示したことが分かる．

時系列性を考慮することで推測精度を向上させられるかの検証と時系列の順序性の検証した結果，今回の実験条件

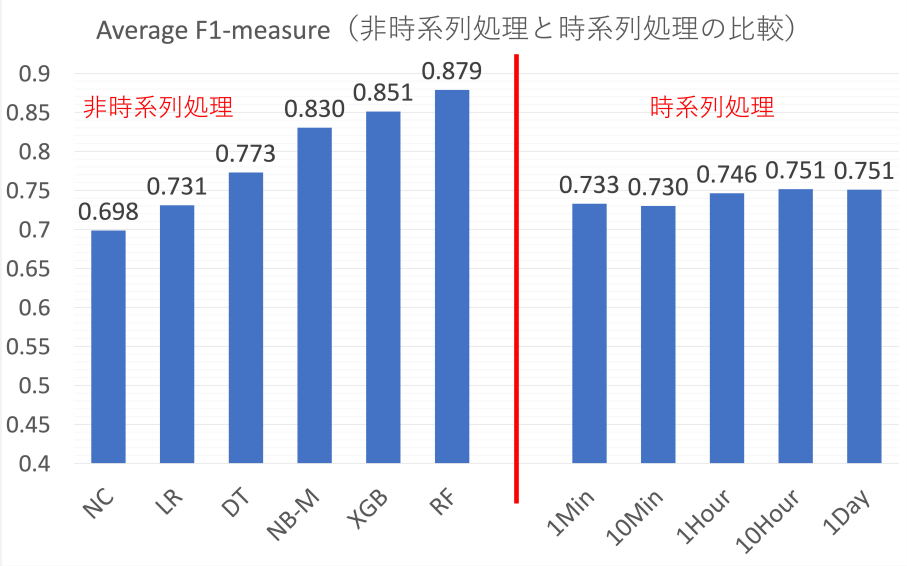

図 11　時系列処理と非時系列処理の比較

Fig. 11　Comparison of time series and non-time series processing results.

においては精度を向上させることができなかった．しかしながら，予測対象店舗によっては，直前の行動の影響を受けるケースや，分析するデータ量に大きな違いが出るケースなども考えられるため，予測対象店舗を含む各種分析条件を変化させることで，今回とは異なる結果を示す可能性もある．したがって，今後も様々な条件での評価実験および考察を行いたいと考えている．

4.4　提案手法の評価および採用したデータに対する考察

本研究では，実空間においてユーザが行動した周辺エリアの店舗カテゴリを考慮した潜在的興味分析として，過去にある特定店舗への訪問履歴がないユーザが，今後この店舗を訪問するかどうかを推定する手法を提案しており，本節でその評価を行った．今回の実験のエリアサイズの比較（図 5，図 6，図 11，図 9）では，多くのアルゴリズムにおいて考慮するエリアサイズを大きくすると予測精度が向上している．このことから周辺エリアの店舗カテゴリを考慮することは有効ではないかと考える．しかしながら現状は対象店舗訪問予測であるため，実際のユーザ評価で提案手法をもとに推薦された広告をユーザに提示した場合，広告の店舗に訪れるのか検証が必要である．

採用したデータに対する考察としては，明示的に投稿したかそうでないかの違いが興味推定に与える影響は大きいのではないかと考える．非時系列処理アルゴリズム・時系列処理アルゴリズムともにユーザ行動ログデータの方が良い結果を示していた．明示的に投稿する場合とそうでない場合を比べると，明示的に投稿する方が記録回数が少なくなるため，その記録回数の差が与える影響は大きいと考える．また潜在的興味を推定する場合，ふだんは自ら記録しないスポット情報を考慮に入れることでより精度の高い予測ができたのではないかと考えるからである．

5.　おわりに

本稿では実空間のユーザ行動分析に基づく潜在的興味分析方式について提案し，2 種類の分析データおよび種々の学習アルゴリズムによる評価実験および考察を行った．

特徴ベクトル作成時の周辺スポットを考慮するエリアサイズに関しては，50 メートル四方から 300 メートル四方のうち，より広い 300 メートル四方のエリアサイズが良いという結果が得られた．また，ジオタグ付きツイートデータにおいてユーザの潜在的興味推定に必要な分析対象期間については，1 週間から 3 カ月の期間のうち，今回の実験では 3 カ月が最も結果が良かった．今回の実験では，条件によって分析に必要なデータ量を十分に確保することが難しいケースもあり，分析データを確保しやすい広いエリアサイズや，より長い分析対象期間において良い結果を得ることができた可能性がある．今後は確保できるデータ量と予測性能の関係も含めて調査を続けるつもりである．

時系列処理のアルゴリズム比較では，AttBiLSTM が最も良かったため時系列処理アルゴリズムによるエリアサイズの比較では AttBiLSTM を使用した．エリアサイズの比較においては多くの情報を考慮するために広いエリアの方が良く，Window サイズは 1 日が最も良いという結果が得られた．非時系列処理アルゴリズムと時系列処理アルゴリズムを比較した結果，本実験条件においては非時系列処理アルゴリズムである Random Forest が最も良い結果を示した．

非時系列処理，時系列処理ともに今回の実験条件で最も広いエリアサイズ，最も長い期間，最も長い Window サイズが良いという結果が得られたためさらに広いエリアサイ

ズ，さらに長い期間，さらに長い Window サイズで検証を行う必要があると考える．

なお，提案手法による潜在的興味分析（対象店舗への訪問予測）の結果は，今回の実験条件に対する結果であり，予測対象店舗や各種実験条件によっては異なる結果を示す可能性もある．

本研究は将来的には実空間にて活動中のユーザに対する実店舗の広告推薦が可能なシステムの開発を目指しているが，このシステムの汎用性を高めるためにも，予測対象店舗や各種実験条件の変更も行いながら，今後さらに調査を続けたいと考えている．

謝辞　本研究の一部は，科学研究費（課題番号：19H04118，20H04293，20H00584，19K12240）および京都産業大学先端科学技術研究所（ヒューマン・マシン・データ共生科学研究センター）共同研究プロジェクト（M2001）の助成を受けたものである．ここに記して謝意を表す．

参考文献

[1] 株式会社サイバー・コミュニケーションズ（CCI），株式会社 D2C，株式会社電通，株式会社電通デジタル：「2020 年日本の広告費 インターネット広告媒体費 詳細分析」(2021)，入手先 ⟨https://www.dentsu.co.jp/news/release/2021/0310-010348.html⟩．

[2] Li, K. and Du, T.C.: Building a targeted mobile advertising system for location-based services, *Decision Support Systems*, Vol.54, No.1, pp.1–8, DOI: 10.1016/j.dss.2012.02.002 (2012).

[3] Lian, S., Cha, T. and Xu, Y.: Enhancing geotargeting with temporal targeting, behavioral targeting and promotion for comprehensive contextual targeting, *Decision Support Systems*, Vol.117, pp.28–37, DOI: 10.1016/j.dss.2018.12.004 (2019).

[4] 内野英治，森田博彦，下野雅芳：Web 広告動的配信システムへのマルコフモデルと kMER の応用，日本知能情報ファジィ学会ファジィシステムシンポジウム講演論文集第 22 回ファジィシステムシンポジウム，日本知能情報ファジィ学会，pp.61–62, DOI: 10.14864/fss.22.0.17.0 (2006).

[5] 小河真之，原田史子，島川博光ほか：消費者の情報探索行動に着目した広告の内容と表示の個別化，研究報告データベースシステム（DBS），Vol.2010, No.17, pp.1–8 (2010).

[6] 山口由莉子，Siriaraya, P.，森下民平，稲垣陽一，中本レン，張　建偉，青井順一，中島伸介：Web 広告推薦のための長期的・短期的興味を考慮したユーザの潜在的興味分析方式，第 10 回データ工学と情報マネジメントに関するフォーラム（DEIM Forum 2018），B2-3 (2018).

[7] Ren, K., Qin, J., Fang, Y., Zhang, W., Zheng, L., Bian, W., Zhou, G., Xu, J., Yu, Y., Zhu, X., et al.: Lifelong sequential modeling with personalized memorization for user response prediction, *Proc. 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.565–574 (2019).

[8] 倉島　健，岩田具治，星出高秀，高屋典子，藤村　考：行動範囲と興味の同時推定モデルによる地域情報推薦，情報処理学会論文誌データベース（TOD），Vol.6, No.2, pp.30–41 (2013).

[9] Song, C., Wen, J. and Li, S.: Personalized POI recommendation based on check-in data and geographical-regional influence, *Proc. 3rd International Conference on Machine Learning and Soft Computing*, pp.128–133 (2019).

[10] Gao, Q., Zhou, F., Trajcevski, G., Zhang, K., Zhong, T. and Zhang, F.: Predicting human mobility via variational attention, *The World Wide Web Conference*, pp.2750–2756 (2019).

[11] Li, X., Han, D., He, J., Liao, L. and Wang, M.: Next and next new POI recommendation via latent behavior pattern inference, *ACM Trans. Information Systems* (*TOIS*), Vol.37, No.4, pp.1–28, DOI: 10.1145/3354187 (2019).

[12] Safavian, S.R. and Landgrebe, D.: A survey of decision tree classifier methodology, *IEEE Trans. Systems, Man, and Cybernetics*, Vol.21, No.3, pp.660–674, DOI: 10.1109/21.97458 (1991).

[13] Peng, C.-Y.J., Lee, K.L. and Ingersoll, G.M.: An introduction to logistic regression analysis and reporting, *The Journal of Educational Research*, Vol.96, No.1, pp.3–14, DOI: 10.1080/00220670209598786 (2002).

[14] McIntyre, R.M. and Blashfield, R.K.: A nearest-centroid technique for evaluating the minimum-variance clustering procedure, *Multivariate Behavioral Research*, Vol.15, No.2, pp.225–238, DOI: 10.1207/s15327906mbr1502_7 (1980).

[15] Rish, I. et al.: An empirical study of the naive Bayes classifier, *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, Vol.3, No.22, pp.41–46 (2001).

[16] Ho, T.K.: Random decision forests, *Proc. 3rd International Conference on Document Analysis and Recognition*, Vol.1, pp.278–282, IEEE (1995).

[17] Chen, T. and Guestrin, C.: Xgboost: A scalable tree boosting system, *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data mining*, pp.785–794 (2016).

[18] Fukunaga, K. and Narendra, P.M.: A branch and bound algorithm for computing k-nearest neighbors, *IEEE Trans. Computers*, Vol.100, No.7, pp.750–753, DOI: 10.1109/T-C.1975.224297 (1975).

[19] Gardner, M.W. and Dorling, S.: Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences, *Atmospheric environment*, Vol.32, No.14-15, pp.2627–2636, DOI: 10.1016/S1352-2310(97)00447-0 (1998).

[20] McCallum, A., Nigam, K., et al.: A comparison of event models for naive bayes text classification, *AAAI-98 Workshop on Learning for Text Categorization*, Vol.752, No.1, pp.41–48, Citeseer (1998).

[21] Crammer, K., Dekel, O., Keshet, J., Shalev-Shwartz, S. and Singer, Y.: Online passive aggressive algorithms, *Journal of Machine Learning Research*, Vol.7, No.19, pp.551–585 (2006).

[22] Freund, Y. and Schapire, R.E.: Large margin classification using the perceptron algorithm, *Machine Learning*, Vol.37, No.3, pp.277–296, DOI: 10.1023/A:1007662407062 (1999).

[23] Hoerl, A.E. and Kennard, R.W.: Ridge regression: Biased estimation for nonorthogonal problems, *Technometrics*, Vol.12, No.1, pp.55–67, DOI: 10.1080/00401706.1970.10488634 (1970).

[24] Cortes, C. and Vapnik, V.: Support vector machine, *Machine Learning*, Vol.20, No.3, pp.273–297 (1995).

[25] Hochreiter, S. and Schmidhuber, J.: Long short-term memory, *Neural computation*, Vol.9, No.8, pp.1735–1780, DOI: 10.1162/neco.1997.9.8.1735 (1997).

[26] Graves, A. and Schmidhuber, J.: Framewise phoneme

classification with bidirectional LSTM and other neural network architectures, *Neural Networks*, Vol.18, No.5-6, pp.602–610, DOI: 10.1016/j.neunet.2005.06.042 (2005).
[27] Li, L., Liu, Y. and Zhou, A.: Hierarchical attention based position-aware network for aspect-level sentiment analysis, *Proc. 22nd Conference on Computational Natural Language Learning*, pp.181–189 (2018).

大村 貴信

京都産業大学大学院先端情報学研究科博士前期課程在学中．2021 年京都産業大学コンピュータ理工学部ネットワークメディア学科卒業．Web 広告推薦に関する研究に従事．日本データベース学会学生会員．

鈴木 健太

京都産業大学大学院先端情報学研究科博士前期課程在学中．2020 年京都産業大学コンピュータ理工学部ネットワークメディア学科卒業．幸福度向上に関する研究に従事．

パノット シリアーラヤ

2013 年に英国のケント大学で電子工学の博士号を取得．2014 年から 2017 年までデルフト工科大学研究員，2017 年から 2019 年までに京都産業大学研究員を経て，2019 年より京都工芸繊維大学助教，現在に至る．ヒューマンコンピュータインタラクション（HCI），ゲーミフィケーション，高齢者向けのシステムデザイン，レコメンダーシステム等の研究に従事．

栗 達

2020 年北海道大学大学院情報科学研究科博士後期課程修了．北海道大学大学院情報科学研究院専門研究員を経て，2021 年より京都産業大学情報理工学部研究員．主に自然言語処理，機械学習，データマイニングおよび感情分析の研究に従事．日本人工知能学会，AAAI，IEEE Computer Society 各会員．

河合 由起子　（正会員）

2001 年奈良先端科学技術大学院大学情報科学研究科博士後期課程修了．同年，独立行政法人通信総合研究所（現，国立研究開発法人情報通信研究機構），2006 年京都産業大学理学部講師を経て，2018 年より京都産業大学情報理工学部教授，大阪大学サイバーメディアセンター特任教授（常勤），現在に至る．博士（工学）．Web マイニング，時空間分析，情報推薦の研究に従事．電子情報通信学会，日本データベース学会各会員．

中島 伸介　（正会員）

1997 年神戸大学大学院自然科学研究科博士前期課程修了．2004 年京都大学大学院情報学研究科博士後期課程修了．博士（情報学）．情報通信研究機構専攻研究員，奈良先端科学技術大学院大学助教，京都産業大学コンピュータ理工学部准教授および教授を経て，2015 年より京都産業大学情報理工学部教授．主に Web マイニングおよび情報推薦の研究に従事．電子情報通信学会，日本データベース学会，環境システム計測制御学会，ACM，IEEE Computer Society 各会員．

（担当編集委員　佐藤 翔）