DEIM2023 5c-4-4

ユーザ行動分析のエリアサイズが訪問確率予測精度に与える影響

大村　貴信† Panote Siriaraya†† 栗　達††† 田中　克己††††
河合由起子††††† 中島　伸介†††††

† 京都産業大学大学院 先端情報学研究科　〒 603–8555 京都府京都市北区上賀茂本山
†† 京都工芸繊維大学 情報工学・人間科学系　〒 606–8585 京都市左京区松ヶ崎橋上町
††† 福岡大学　工学部　〒 814–0180 福岡県福岡市城南区七隈 8 丁目 19 番 1 号
†††† 福知山大学　情報学部　〒 620-0886 京都府福知山市字堀 3370
††††† 京都産業大学 情報理工学部　〒 603–8555 京都府京都市北区上賀茂本山
E-mail: †i2186023@cc.kyoto-su.ac.jp, ††spanote@kit.ac.jp, †††lida@fukuoka-u.ac.jp,
††††tanaka-katsumi@fukuchiyama.ac.jp, †††††{kawai,nakajima}@cc.kyoto-su.ac.jp

あらまし　インターネット向けの Web 広告サービスは年々増加傾向にある．しかし，現在主流である検索クエリや閲覧履歴とのキーワードマッチングに基づく Web 広告推薦手法では，潜在的興味を持つユーザに効果的に Web 広告を推薦することは難しい．これは，ユーザが既に興味を持っているキーワードを用いた明示的な分析であるためだ．一方，携帯端末の位置情報を利用した広告推薦手法は，実店舗までの距離に基づくものがほとんどである．そこで我々は，実空間ユーザ行動データを用いて潜在的興味を分析することで，より効果的な Web 広告推薦方式を実現することを目的に研究を進めている．提案手法では特徴ベクトル生成時に，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ではなく，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．そこで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が，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．そこで本稿では，ユーザ行動分析のエリアサイズが訪問確率予測精度に与える影響を調査する．具体的に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の行動範囲を抽出する．抽出された行動範囲からユーザが訪問した場所をスポット属性毎にカウントし特徴ベクトルを作成する．この特徴ベクトルとクラス分類手法を用いて学習モデルを生成する．本稿では複数のエリアサイズ，複数のクラス分類手法を用いて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

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] にある．しかし，現在主流であるネット検索や閲覧履歴とのキーワードマッチングに基づく Web 広告推薦手法では，潜在的興味を持つユーザに効果的に Web 広告を推薦することは難しい．これは，ユーザが既に興味を持っているキーワードを用いた明示的な分析であるためだ．一方，携帯端末の位置情報を利用した広告推薦手法は，実店舗までの距離に基づくものがほとんどである．そこで我々は，実空間ユーザ行動データを用いて潜在的興味を分析することで，より効果的な Web 広告推薦方式を実現することを目的に研究を進めている．提案手法では特徴ベクトル生成時に，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ではなく，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．そこで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が，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．そこで本稿では，ユーザ行動分析のエリアサイズが訪問確率予測精度に与える影響を調査する．

具体的に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の行動範囲を抽出する．抽出された行動範囲からユーザが訪問した場所をスポット属性毎にカウントし特徴ベクトルを作成する．この特徴ベクトルとクラス分類手法を用いて学習モデルを生成する．本稿では複数のエリアサイズ，複数のクラス分類手法を用いて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] と言われる手法の

一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に基づいて広告をパーソナライズするターゲット方法である．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繋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し，実店舗から一定距離圏内にいるユーザに向けて広告を推薦するものである．これに対して，提案手法ではユーザの位置情報に加えて実空間での行動に対する意味的な分析を併せて行うことでユーザの好みや特性を考慮した広告を推薦するものであり，独自性・新規性は高いと考えている．

本稿の構成は以下の通りである．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 章では実験計画について述べる．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 [6] 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 [7] 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 [8] 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) を予測及び推薦する研究も盛んに行われている．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 [9]．これは食べログ (注1) の店舗への訪問履歴（レビュー履歴）と Flickr(注2) の写真のジオタグ情報を用いて行動履歴を再現し，人間が行動する時の特徴を用いた 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 [10]．これは Gowalla(注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．他には Location-Based Social Networks (LBSNs) のチェックイン記録が疎であるため，POI 予測及び推薦するこ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 [11] 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 [12] もある．これらの研究は Foursquare(注4) や Gowalla のデータを用いて POI 推薦を行っている．関連研究では人間が行動する時の特徴の使用や 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．

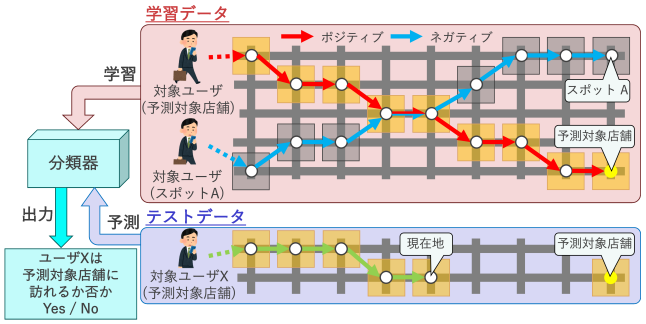


図 1: 提案手法のシステム概要

図 2: 提案システムの推薦の例

3. スポット属性を考慮した実空間のユーザ行動分析に基づく潜在的興味推定方式

本章では，スポット属性を考慮した実空間のユーザ行動分析に基づく潜在的興味推定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

(注1)：https://tabelog.com/
(注2)：https://www.flickr.com/
(注3)：https://go.gowalla.com/
(注4)：https://foursquare.com/

表 1: 本研究で採用したデータの基本統計情報

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

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等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．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 データ収集とポジティブ・ネガティブ分類

本研究では 2 種類のデータを使用する．一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．2019 年のある 1ヶ月間に収集された，約 2 億件のデータであり，該当する予測対象店舗は，幅広い年齢層が訪れる日用品・食品等も扱う小売業態の 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無い様，幅広いユーザが訪れるような店舗を選定した．もう一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．数値は全てユーザ行動ログデータは 1ヶ月間，ジオタグ付きツイートデータは 4 年間のものである．

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

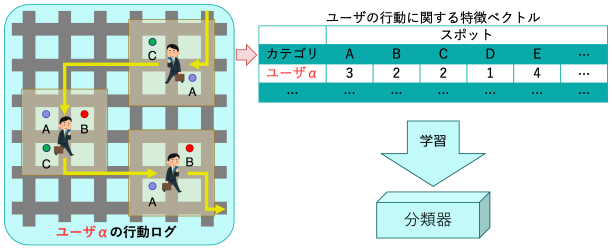


図 3: ユーザ行動特徴ベクトルの抽出

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全て “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，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．これら候補のうち，位置情報の記録回数が不自然に多い場合や少ない場合には，学習データから排除し，ジオタグ付きツイートデータでは 20 100 ツイートであるユーザデータ，ユーザ行動ログデータでは 90〜4,000 回であるユーザデータを採用した．

3.2 ユーザ行動の特徴抽出

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(注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毎にとりまとめ，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) 毎にカウントし，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

(注5)：https://www.openstreetmap.org/

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，さらに単位時間毎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推定手法

予測対象店舗への潜在的興味の推定手法としては，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時，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. ユーザ行動分析のエリアサイズが訪問確率予測精度に与える影響の調査

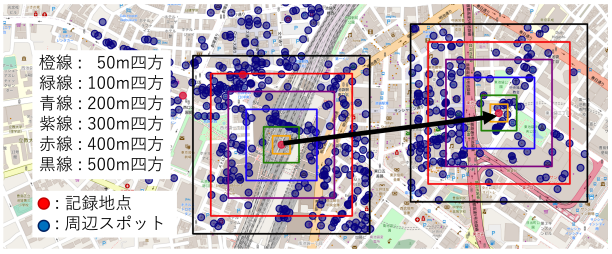


図 4: ユーザの潜在的興味として青のスポットを検出

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推定手法として，過去にある特定店舗への訪問履歴が無いユーザが，今後この店舗を訪問するか否かを推定する手法を提案している．提案手法では特徴ベクトル生成時に，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．そこで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が，このとき考慮するエリアサイズの検討は重要である．図 4 のように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．そのため，本稿では，ユーザ行動分析のエリアサイズが訪問確率予測精度に与える影響を調査した．また同時に対象店舗訪問予測を行う上で，最も適したアルゴリズムの検証も行なったので報告する．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．ユーザ行動ログデータにおいてポジティブ・ネガティブユーザの数が同数であるのはランダムに選択したユーザデータに対して OpenStreetMap データを付与したためである．ユーザ行動ログデータがジオタグ付きツイートデータに比べデータ量が多く，OpenStreetMap のスポットを紐づける作業にかなりの時間を要するため，ランダムに選択したユーザデータに対し紐付け作業を行なった．

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

	ジオタグ付きツイートデータ	ユーザ行動ログデータ
ポジティブユーザ	232	200
ネガティブユーザ	597	200
レコード数	38,369	518,818
1 ユーザあたりの平均記録回数	46	1,297
1 ユーザあたりの平均スポット数	6,812	82,850

4.1 非時系列処理アルゴリズムによる潜在的興味推定

本研究ではスポット属性を考慮した実空間のユーザ行動分析に基づく潜在的興味推定手法を提案しており，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いる．しかし本研究で採用したジオタグ付きツイートデータのように連続性が保持できない時間的に疎なデータでは，訪問順序に意味があるかわからない可能性もあるため，本節で非時系列処理アルゴリズムによる興味分析の評価を行う．

周辺スポットを考慮するエリアサイズとしては 500 メートル四方，400 メートル四方，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 6 種類のエリアサイズのデータを用意し，ポジティブユーザ・ネガティブユーザ共に 120 人，合計 240 ユーザの特徴ベクトルをランダムに選択し，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共に 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，Support Vector Machine-rbf (SVM-rbf) [13]，Naïve Bayes-Multinomial (NB-M) [14]，Random Forest (RF) [15]，XGBoost (XGB) [16]，Decision Tree [17]，Logistic Regression [18]，Nearest Centroid [19]，k-Nearest Neighbors [20]，Multilayer Perceptron [21]，Naïve Bayes-Bernoulli [22] ，Passive Aggressive Classifier [23]，Perceptron [24]，Ridge Regression [25]，Support Vector Machine-linear [13] 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 つの計 4 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毎の F 値比較を，図 6 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毎の 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，Support Vector Machine の rbf カーネルの 200 メートル四方が最も結果が良い．図 6 のユーザ行動ログデータの結果では，Random Forest と XGBoost 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の Support Vector Machine の rbf カーネルで 0.772 という結果に比べ，ユーザ行動ログデータは Random Forest と XGBoost で 0.873 と良い結果を

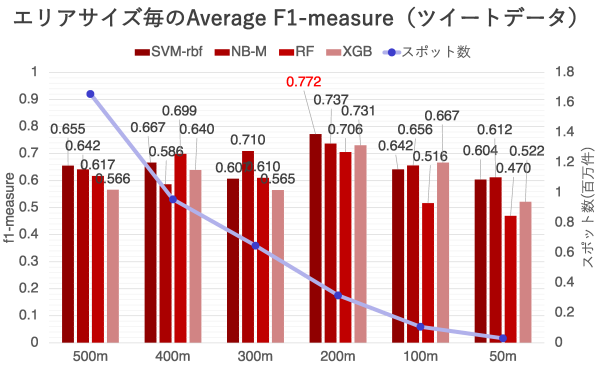


図 5: ジオタグ付きツイートデータのエリアサイズ毎の F 値（非時系列処理）

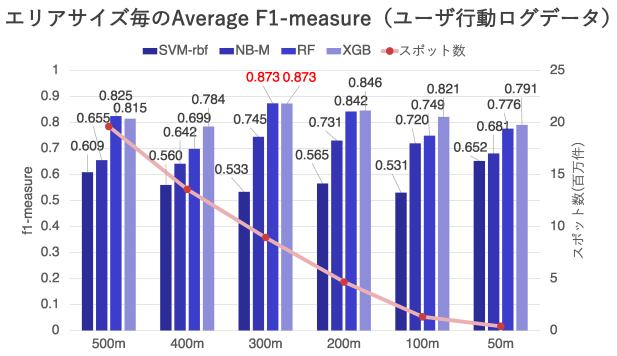


図 6: ユーザ行動ログデータのエリアサイズ毎の F 値（非時系列処理）

示している．この一因としてデータの密度の違いが挙げられる．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 3 万件から 170 万件であるのに対し，ユーザ行動ログデータでは 40 万件から 2000 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言える．このことがユーザ行動ログデータの方が良い結果を示した一因であると考えられる．

エリアサイズの比較として本実験条件においては 200m〜300m 四方が良い結果を示した．500m，400m 四方では移動地点とは関係のないエリアを多く含んだことにより興味の抽出が難しくなり,，50m100m 四方ではデータ量が少なかったため興味を抽出するのが難しかったことが影響していると考える．

4.2 時系列処理アルゴリズムによる潜在的興味推定

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予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

本節では，各種時系列アルゴリズムのうち，提案手法に適用する上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．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 サイズ毎に区切って，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．ここでは，エリアサイズを 500 メートル四方，400 メートル四方，300 メートル四方，Window サイズを 1 日と 10 時間としたデータを用意し，ポジティブユーザ・ネガティブユーザ共に 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) [26]，bidirectional LSTM (Bi-LSTM) [27]，attention-based bidirectional LSTM (AttBiLSTM) [28] 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 7，図 8 に示す．

図 7 に，ジオタグ付きツイートデータの時系列処理アルゴリズムの F 値比較を，図 8 に，同じくユーザ行動ログデータの時系列処理アルゴリズムの F 値比較を示す．図 7，図 8 ともに AttBiLSTM が最も結果が良い．そこで次節のエリアサイズの比較実験では 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，500 メートル四方，400 メートル四方，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 6 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 30 種類のデータを用意し，ポジティブユーザ・ネガティブユーザ共に 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共に 3:1 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毎の 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毎の F 値（時系列処理）を示す．

図 9 では 400 メートル四方の 1 日が最も良く，図 10 では 500 メートル四方の 10 時間が最も結果が良い．また，わずかではあるがジオタグ付きツイートデータ，ユーザ行動ログデータ共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．時間ごとに区切っていることから疎なベクトルとなるためユーザの興味を捉えるために十分なデータ量を確保できる広いエリアの結果が良くなった可能性があると考える．また Window サイズごとに大きな差は見られなかった．これも時間ごとに区切っていることから疎なベクトルとなったことが一因ではないかと考える．

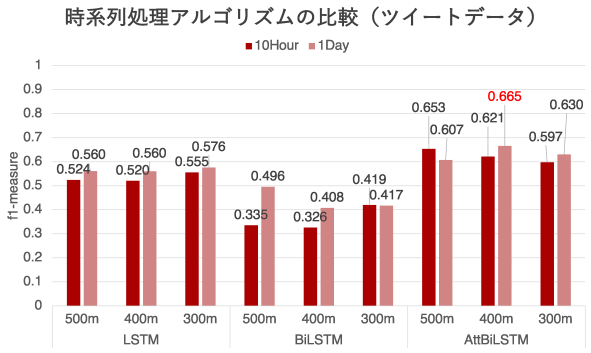


図 7: ジオタグ付きツイートデータの時系列処理アルゴリズムの比較

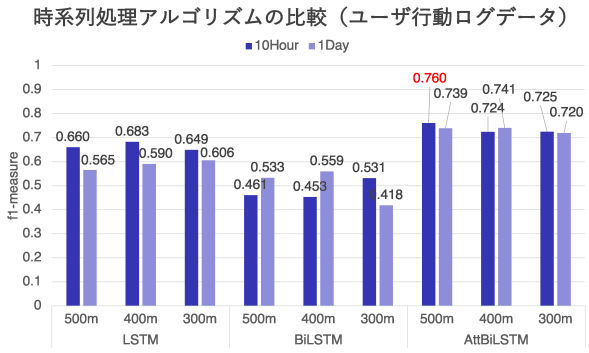


図 8: ユーザ行動ログデータの時系列処理アルゴリズムの比較

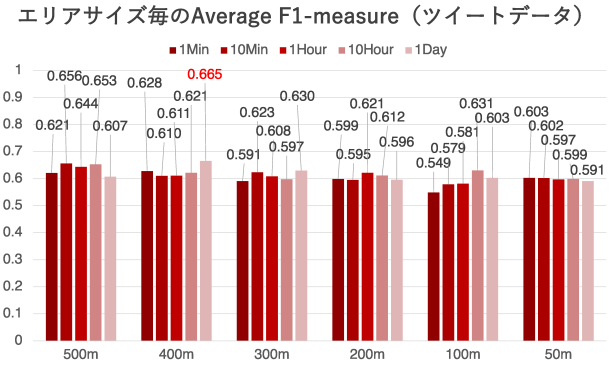


図 9: ジオタグ付きツイートデータのエリアサイズ毎の F 値（時系列処理）

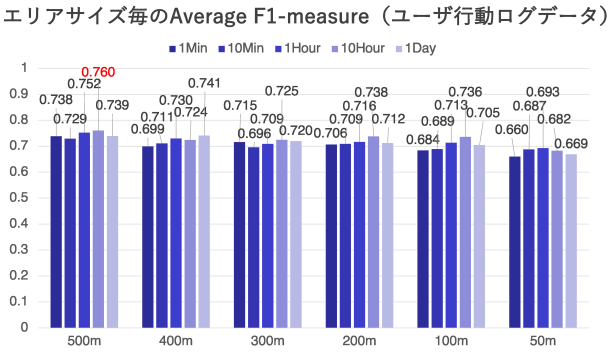


図 10: ユーザ行動ログデータのエリアサイズ毎の F 値（時系列処理）

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推定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 11a にジオタグ付きツイートデータの非時系列処理で最も良い結果を得られた 200 メートル四方，図 11b にジオタグ付きツイートデータの時系列処理で最も良い結果を得られた 400 メートル四方，図 12a にユーザ行動ログデータの非時系列処理で最も良い結果を得られた 300 メートル四方，図 12b にユーザ行動ログデータの時系列処理で最も良い結果を得られた 500 メートル四方の時系列処理アルゴリズムおよび非時系列処理アルゴリズムで比較したグラフを示す．図 11 と図 12 から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の特に Random Forest の最も良い結果を示した．

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．また次元削減を行うことで時系列処理アルゴリズムが良い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．

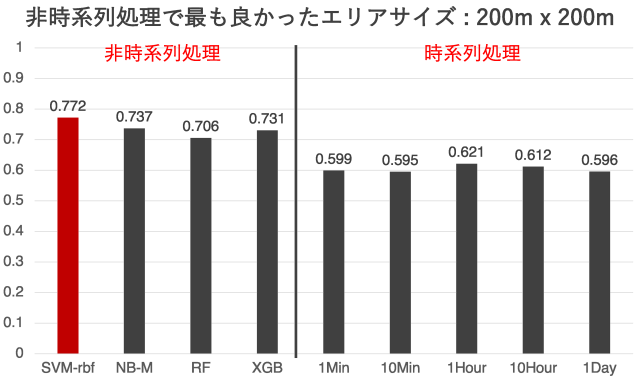
5. おわりに

本稿ではスポット属性を考慮した実空間のユーザ行動分析に基づく潜在的興味推定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．

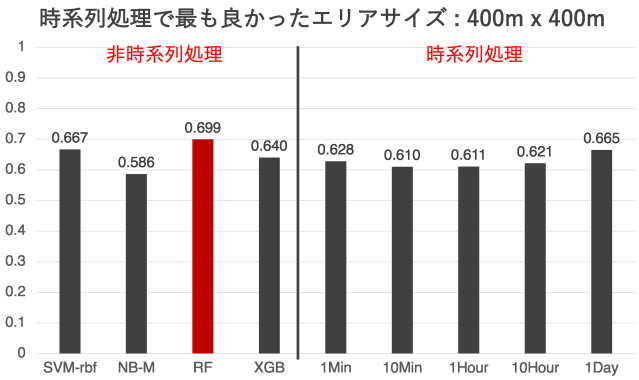
特徴ベクトル作成時の周辺スポットを考慮するエリアサイズに関しては，非時系列処理では 50 メートル四方から 300 メートル四方のうち，200 メートル四方から 300 メートル四方のエリアサイズが良いという結果が得られた．また時系列処理ではデータ量を確保できる広いエリアの方が良いという結果が得られた．

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いという結果が得られた．また時系列データを区切る Window サイズによる大きな差は見られなかった．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．

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

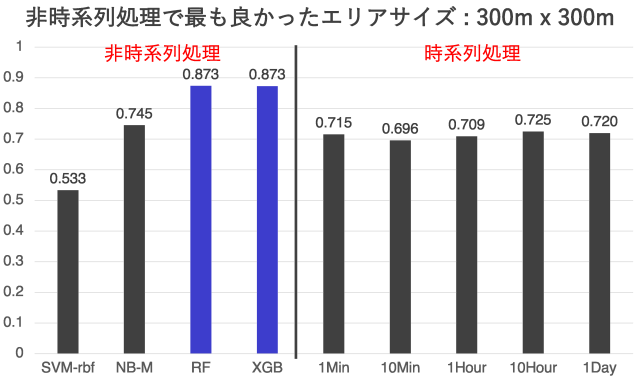


(a) 非時系列処理で最もよかったエリアサイズ:200m x 200m

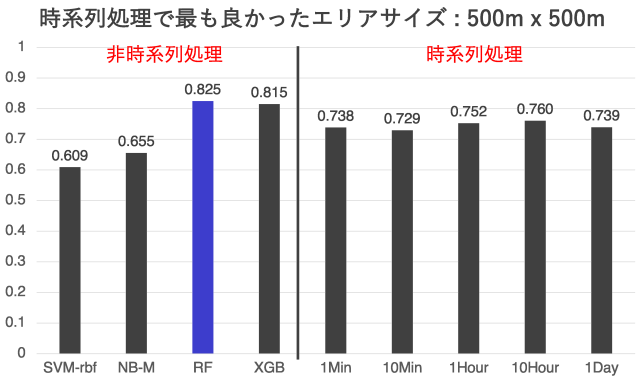


(b) 時系列処理で最もよかったエリアサイズ:400m x 400m

図 11: 時系列処理と非時系列処理の比較 (ツイートデータ)



(a) 非時系列処理で最もよかったエリアサイズ:300m x 300m



(b) 時系列処理で最もよかったエリアサイズ:500m x 500m

図 12: 時系列処理と非時系列処理の比較 (ユーザ行動ログデータ)

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．

文　　献

[1] 株式会社電通.「2021 年 日本の広告費」. https://www.dentsu.co.jp/news/release/2022/0224-010496.html. (Accessed on 11/01/2023).

[2] Kai Li and Timon C. Du. Building a targeted mobile advertising system for location-based services. *Decision Support Systems*, Vol. 54, No. 1, pp. 1–8, 2012. doi: https://https://doi.org/10.1016/j.dss.2012.02.002.

[3] Shaohua Lian, Tingting Cha, and Yunjie Xu. Enhancing geotargeting with temporal targeting, behavioral targeting and promotion for comprehensive contextual targeting. *Decision Support Systems*, Vol. 117, pp. 28–37, 2019. doi: https://https://doi.org/10.1016/j.dss.2018.12.004.

[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, pp. 61–62. 日本知能情報ファジィ学会, 2006. doi: https://doi.org/10.14864/fss.22.0.17.0.

[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] 久松俊道, 外川隆司, 朝日弓未, 生田目崇. Ec サイトにおける購買予兆発見モデルの提案. オペレーションズ・リサーチ: 経営の科学, Vol. 58, No. 2, pp. 93–100, 2013.

[7] 山口由莉子, Panote Siriaraya, 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.

[8] Kan Ren, Jiarui Qin, Yuchen Fang, Weinan Zhang, Lei Zheng, Weijie Bian, Guorui Zhou, Jian Xu, Yong Yu, Xiaoqiang Zhu, et al. Lifelong sequential modeling with personalized memorization for user response prediction. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 565–574, 2019.

[9] 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.

[10] Chuang Song, Junhao Wen, and Shun Li. Personalized poi recommendation based on check-in data and geographical-

regional influence. In *Proceedings of the 3rd International Conference on Machine Learning and Soft Computing*, pp. 128–133, 2019.

[11] Qiang Gao, Fan Zhou, Goce Trajcevski, Kunpeng Zhang, Ting Zhong, and Fengli Zhang. Predicting human mobility via variational attention. In *The World Wide Web Conference*, pp. 2750–2756, 2019.

[12] Xin Li, Dongcheng Han, Jing He, Lejian Liao, and Mingzhong Wang. Next and next new poi recommendation via latent behavior pattern inference. *ACM Transactions on Information Systems (TOIS)*, Vol. 37, No. 4, pp. 1–28, 2019. doi: https://doi.org/10.1145/3354187.

[13] Corinna Cortes and Vladimir Vapnik. Support vector machine. *Machine learning*, Vol. 20, No. 3, pp. 273–297, 1995.

[14] Irina Rish, et al. An empirical study of the naive bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, Vol. 3, pp. 41–46, 2001.

[15] Tin Kam Ho. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, Vol. 1, pp. 278–282. IEEE, 1995.

[16] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, 2016.

[17] S Rasoul Safavian and David Landgrebe. A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics*, Vol. 21, No. 3, pp. 660–674, 1991. doi: https://doi.org/10.1109/21.97458.

[18] Chao-Ying Joanne Peng, Kuk Lida Lee, and Gary M Ingersoll. An introduction to logistic regression analysis and reporting. *The journal of educational research*, Vol. 96, No. 1, pp. 3–14, 2002. doi: https://doi.org/10.1080/00220670209598786.

[19] Robert M McIntyre and Roger K Blashfield. A nearest-centroid technique for evaluating the minimum-variance clustering procedure. *Multivariate Behavioral Research*, Vol. 15, No. 2, pp. 225–238, 1980. doi: https://doi.org/10.1207/s15327906mbr1502_7.

[20] Keinosuke Fukunaga and Patrenahalli M. Narendra. A branch and bound algorithm for computing k-nearest neighbors. *IEEE transactions on computers*, Vol. 100, No. 7, pp. 750–753, 1975. doi: https://doi.org/10.1109/T-C.1975.224297.

[21] Matt W Gardner and SR Dorling. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric environment*, Vol. 32, No. 14-15, pp. 2627–2636, 1998. doi: https://doi.org/10.1016/S1352-2310(97)00447-0.

[22] Andrew McCallum, Kamal Nigam, et al. A comparison of event models for naive bayes text classification. In *AAAI-98 workshop on learning for text categorization*, Vol. 752, pp. 41–48. Citeseer, 1998.

[23] Koby Crammer, Ofer Dekel, Joseph Keshet, Shai Shalev-Shwartz, and Yoram Singer. Online passive aggressive algorithms. 2006.

[24] Yoav Freund and Robert E Schapire. Large margin classification using the perceptron algorithm. *Machine learning*, Vol. 37, No. 3, pp. 277–296, 1999. doi: https://doi.org/10.1023/A:1007662407062.

[25] Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, Vol. 12, No. 1, pp. 55–67, 1970. doi: https://doi.org/10.1080/00401706.1970.10488634.

[26] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, Vol. 9, No. 8, pp. 1735–1780, 1997. doi: https://doi.org/10.1162/neco.1997.9.8.1735.

[27] Alex Graves and Jürgen Schmidhuber. Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural networks*, Vol. 18, No. 5-6, pp. 602–610, 2005. doi: https://doi.org/10.1016/j.neunet.2005.06.042.

[28] Lishuang Li, Yang Liu, and AnQiao Zhou. Hierarchical attention based position-aware network for aspect-level sentiment analysis. In *Proceedings of the 22nd conference on computational natural language learning*, pp. 181–189, 2018.