

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 潜在的興味分析の検証

大村 貴信[†] 鈴木 健太[†] Panote Siriaraya^{††} 栗 達^{†††}

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

^{†††} 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,i2086060}@cc.kyoto-su.ac.jp, ^{††}spanote@kit.ac.jp,

^{†††}{lida,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、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分析手法を提案し、検証を行う。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つき、一定時間内に行動したユーザの行動範囲を抽出する。OpenStreetMap (OSM) データを用いてこの行動範囲のスポット属性分析を行うことで、実空間行動履歴の特徴を抽出し、学習モデルを生成する。先行研究では蓄積されたデータを学習データとテストデータに分け、様々な条件で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してきた。そこで本稿では提案手法の有効性を検証するために実ユーザに対して行う、アンケート調査について説明する。

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、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分析手法を提案し、検証を行う。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無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、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つき、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。先行研究 [2] では、蓄積されたデータを学習データとテストデータに分け、様々な条件で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してきた。そこで本稿では提案手法の有効性を検証するために実ユーザに対して行う、アンケート調査について説明する。

本研究は一般にジオターゲティング [3] [4] 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 [5] 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 [6]、閲覧行動パターンを考慮した購買兆の発見モデルを提案している研究 [7] 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 [8] 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 [9] 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 [10]。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 [11]。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 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 [12] 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 [13] 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

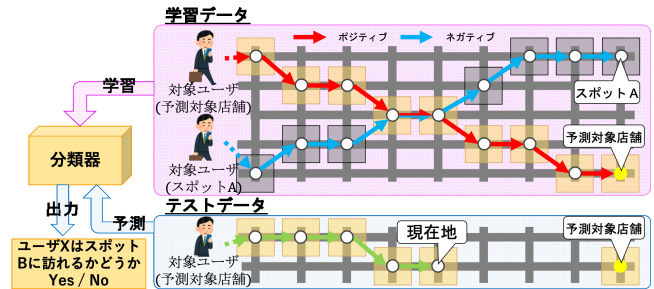


図 1 提案手法のシステム概要



図 2 提案システムの推薦の例

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。

3. 実空間行動範囲の店舗属性分析と潜在的興味推定に基づく広告推薦方式

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分析の概要を解説し、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

(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

		ツイートデータ
期間		4 年間
ユーザ数		3,414
レコード数		6,269,171
1 人あたりの記録回数	平均	1,836
	最大	44,091
	中央	1018
	最小	1

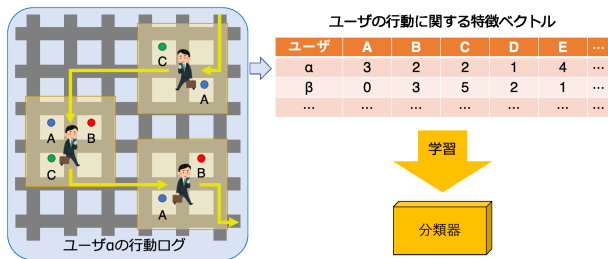


図 3 ユーザ行動特徴ベクトルの抽出

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

本研究ではジオタグ付きツイートデータを使用する。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 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小売業態のチェーン 147 店舗を使用した。表 1 に基本統計情報を示す。

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。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” という内容から判断した。本研究で使っ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。

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、ジオタグ付きツイート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。

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分析の方法としては、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. 提案手法に対する評価

本研究では、ユーザの潜在的興味を抽出し、それに合わせた広告を推薦するための手法を提案し、先行研究では、その最適な条件を調査するために、蓄積されたデータを用いてアルゴリズム・分析対象エリアサイズを比較してきた。そこで本稿では提案手法の有効性を検証するため、実ユーザに対して行うアンケート調査を行った。本章ではその結果を報告する。

4.1 調査内容

アンケート調査にはクラウドワークス^(注6)を使用した。具体的には、以下の手順で回答してもらう。

(1) Google Street View(GSV) で歩いている映像 (1 分程度) を視聴。

(2) 上記映像上のルート上に、スポット (複数回) を提示。

(3) 映像の視聴後、ルート上のスポットとは異なる 5 つの業態の店舗に関して、広告が提示された時にクリックしたいと思うかどうかを「強く思う」「やや思う」「どちらでもない」「やや思わない」「全く思わない」の 5 段階で回答。

これを全部で 5 つのルートに対して回答してもらう。アンケートから得られた予測対象店舗の広告をクリックしたいと思うルートのランキングと提案手法・従来手法の予測対象店舗訪問確率のランキングを作成し、nDCG で評価を行う。

4.2 調査結果

アンケートは 50 人に回答していただいた。この中で映像を視聴したか確認の質問に誤答していた 8 名を除き、42 名の結

(注5) : <https://www.openstreetmap.org/>

(注6) : <https://crowdworks.jp/>

店舗・ルートごとの5段階評価の平均

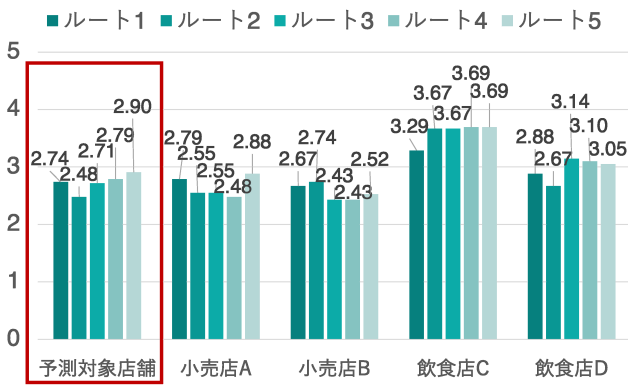


図4 アンケート調査結果

表2 予測対象店舗のルートごとの5段階評価

	ルート1	ルート2	ルート3	ルート4	ルート5
評価平均	2.738	2.476	2.714	2.786	2.905

果を集計した。「強く思う」「やや思う」「どちらでもない」「やや思わない」「全く思わない」を順に5点、4点、3点、2点、1点とする。

図4にアンケート結果を示す。また予測対象店舗に限定した結果を表2に示す。小売店と飲食店をを比較すると飲食店の方が提示された広告を「クリックしたいと思う」という回答が多かった。これは小売店と飲食店への興味の持ち方が異なっていたのではないかと考える。しかし、予測対象店舗の広告をクリックしたいと思うルートの評価には差が出ており、順にルート5、ルート4、ルート1、ルート3、ルート2という結果が得られた。これを正解データとして扱い、後述でnDCGを求める。

4.3 提案手法・従来手法の予測対象店舗訪問確率

本節では提案手法と従来手法を用いてアンケートで提示した5つのルートの予測対象店舗訪問確率を求める。まず、提案手法を用いて分類器を作成し、各ルートの予測対象店舗訪問確率を求める。先行研究[2]において分類器作成の最適条件を調査を行い、特徴ベクトル作成時の考慮するエリアサイズを5種類、アルゴリズムを14種類で比較した結果、エリアサイズは300m四方、アルゴリズムはXGBoost[14]が最も良い結果を得られた。そのため上記条件で分類器を作成する。各ルートを歩いたと仮定した特徴ベクトルを作成し、これをテストデータとして分類器に与え、予測対象店舗訪問確率を求める。

続いて従来手法を用いて分類器を作成し、各ルートの予測対象店舗訪問確率を求める。従来手法は投稿されたツイートに含まれるスポット名を使用して特徴ベクトルを作成する。例えば「I'm at starbucks.」というツイートが投稿されていた場合、提案手法はツイートの投稿場所周辺のスポットのカテゴリ情報を特徴ベクトルに使用するのに対し、従来手法はstarbucksを用いて特徴ベクトルを作成する。このようにして作成した特徴ベクトルを用いて提案手法と同様に14種類のアルゴリズムで比較した結果、Random Forest[15]が最も良い結果を得られた。

表3 提案手法・従来手法の予測対象店舗訪問確率

		ルート1	ルート2	ルート3	ルート4	ルート5
評価	提案手法	0.315	0.167	0.092	0.068	0.551
平均	従来手法	0.870	0.820	0.790	0.920	0.770

表4 nDCGによる評価

	提案手法	従来手法
nDCG@5	0.921	0.875

そのため上記条件で分類器を作成する。各ルートを歩いたと仮定した特徴ベクトルを作成し、これをテストデータとして分類器に与え、予測対象店舗訪問確率を求める。

得られた結果を表3に示す。提案手法は順にルート5、ルート1、ルート2、ルート3、ルート4という結果が、従来手法は順にルート4、ルート1、ルート2、ルート3、ルート5という結果がそれぞれ得られた。

4.4 評価

nDCG[16]で評価を行う。nDCGは式(1)で、DCGは式(2)で表される。

$$DCG = r_1 + \sum_{i=2}^N \left(\frac{r_i}{\log_2 i} \right) \quad (1)$$

$$nDCG = \frac{DCG}{DCG_{answer}} \quad (2)$$

r_i はランキングの*i*番目の要素を表し、 N は評価に用いる要素数を表す。今回は要素数を5とし、順位の利得スコアのみで評価を行う。

アンケート結果による予測対象店舗の広告をクリックしたいと思うルートのランキングと提案手法・従来手法の予測対象店舗訪問確率のランキング、nDCGの結果を表4に示す。本実験条件においては提案手法がnDCGにおいてより良い結果を示したことを確認できた。しかし、アンケート結果において小売店より飲食店の方が結果が良かったことやアンケート結果においてルートごとに大きな差が見られなかったことがあるため、さらなる調査が必要であると考えている。

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分析について提案し、提案手法の有効性を検証するため実ユーザに対してアンケート調査を行い、得られた結果と提案手法、従来手法を比較することで本実験条件においては提案手法の方がより良い結果を示した。しかし、アンケート結果において小売店より飲食店の方が結果が良かったことやアンケート結果においてルートごとに大きな差が見られなかったことがあるため、さらなる調査が必要であると考えている。またルート数を増やすことや被験者数を増やすことも考えている。

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」.
- [2] 大村貴信, 鈴木健太, Panote Siriaraya, 栗達, 河合由起子, 中島伸介. 実空間のユーザ行動分析に基づく潜在的興味分析方式. 情報処理学会論文誌データベース, Vol. 92, .
- [3] 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://doi.org/10.1016/j.dss.2012.02.002>.
- [4] 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://doi.org/10.1016/j.dss.2018.12.004>.
- [5] 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>.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] 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>.
- [14] 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.
- [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] Kalervo Järvelin and Jaana Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Trans. Inf. Syst.*, Vol. 20, No. 4, p. 422–446, oct 2002.