

反復性のある日常行動の影響を考慮したPOI推薦システムの一評価法

榎木 佑真[†] 岡本 一志[†]

[†] 電気通信大学 〒182-0021 東京都調布市調布ヶ丘 1-5-1

E-mail: †{y.nagi,kazushi}@uec.ac.jp

あらまし POI 推薦システムの研究では、POI へのチェックイン (C/I: Check-ins) ログが評価に用いられる。C/I ログには反復性のある日常行動が多く含まれており、先行研究ではシステムの性能を公平に比較できていない可能性がある。本研究では、反復性のある日常行動を C/I ログから除外する手法を提案し、それらの行動の除外が推薦精度・多様性・新規性に与える影響を検証する。実験結果として、反復性のある日常行動を一定数除外した際に多くの POI 推薦システムで精度が低下する一方、GRU4Rec は精度を比較的維持できる傾向があり、反復性のある日常行動がシステムの精度比較に影響を与えていることを確認している。

キーワード 情報推薦, 観光, 評価・データセット, Point of Interest, Location-based Social Network

1 はじめに

近年の観光トレンドの変化により、少人数での旅行割合の増加や分散型旅行の推進など [1], 有名観光地への物見遊山に限らない観光行動の多様化が進んでおり、観光行動の意思決定支援の重要性が高まっていると考える。観光行動の意思決定の例としては、“次にどのスポットに行くか”, “どのようなルートで周遊するか” などが挙げられる。

POI (Point of Interest) 推薦システムは、ユーザの嗜好や位置情報などに基づき、観光施設や飲食店などの地理空間上のスポット (POI) をユーザに推薦するアプリケーションであり、観光行動の意思決定支援に貢献できる。また、POI へのチェックイン (C/I: Check-In) 行動を共有できるサービスとして Location-based Social Network (LBSN) があり、ユーザが新しい POI を発見し探索するための手段にもなっている [2], [3]。C/I は明示的フィードバックの一種であり、LBSN ではユーザが特定の POI に訪問したことを意味する。POI 推薦システムは LBSN との親和性が高く、LBSN の出現は POI 推薦システムの改善を後押ししている [3]。

Swarm (Foursquare)¹をはじめとする LBSN では、各ユーザがいつ・どの POI に C/I したか記録されており、POI 推薦システムの研究ではこの C/I ログが評価に主に用いられている [2], [4], [5], [6]。先行研究では C/I ログを評価用データセットとする際、不人気 POI の削除といった単純な前処理のみを行っている。しかし、C/I ログには観光行動と特性が異なる、非観光行動の一種である通勤・通学など“反復性のある日常行動”が一定数含まれている。このような行動により、学校・オフィス・駅など特定の POI に C/I が偏ることが予想される。そのため、反復性のある日常行動の存在を考慮せずに POI 推薦システムを評価すると、たとえ高い推薦精度を獲得していてもユーザが頻繁に C/I する POI の予測に成功しているに過ぎない可能性がある。観光行動の意思決定支援を目的とした POI 推薦

システムにおいては、観光スポットなどの POI が反復的に C/I されないことを踏まえると、従来の単純な前処理によるデータセットの適用ではシステムの性能を公平に比較できていないと考える。

本研究では C/I ログから反復性のある日常行動を抽出する手法を提案し、そのような行動を除外することで観光行動と類似した特性の行動のみで C/I ログを再構成し、観光行動の意思決定支援を目的とした POI 推薦システムの公平な性能比較のための評価用データセットの適切な利用方法について検討する。反復性のある日常行動を除外するため、ツアー (各ユーザの 1 日の C/I 系列) に含まれる POI の特徴類似性を考慮し、それら行動をツアー単位で抽出する方法を適用する。次にこの C/I ログを用いて複数の代表的な POI 推薦システムの推薦精度・多様性・新規性の各指標を評価し、反復性のある日常行動の除外がこれら指標に与える影響を検証する。

本研究は我々の先行研究 [7] をベースとし、行動抽出性能の検証方法、POI 推薦システムの評価指標、評価用データセット (エリア) を拡張させたものである。

2 関連研究

2.1 POI 推薦システム

POI 推薦の研究では、例えばユーザの地域に対する嗜好を決定する問題など様々な種類の問題が扱われている [3]。本研究では、ユーザのある瞬間での次の訪問 (C/I) POI を予測する Next-POI 問題 [3] を扱う POI 推薦システムに焦点を当てる。

提案されている多くの POI 推薦システムは協調フィルタリングをベースとし、POI に関するユーザ嗜好を学習している [2], [3]。因子分解・確率モデルを活用したシステムも多い [3]。Cheng らはマルコフ連鎖と行列因子分解の両手法を組み合わせた FPMC (Factorizing Personalized Markov Chains) [8] に、ユーザ移動に関する局所領域 (LR: Localized Region) 制約を組み込んだ FPMC-LR を提案している [4]。Zhao らは skipgram [9] による埋め込みベースシステムとして Geo-Teaser を

1: <https://ja.swarmapp.com/>

表 1 Next-POI 問題を扱うことができる推薦システムの先行研究

| システム | アプローチ | 地理 / 時間 |
|----------------|--------------------------------|---------|
| FPMC-LR [4] | マルコフ連鎖・行列因子分解 | ✓ / - |
| Geo-Teaser [2] | Skip-gram w/ Neg. Sampling [9] | ✓ / ✓ |
| POI2Vec [5] | Word2Vec を改良 | ✓ / ✓ |
| ATST-LSTM [6] | LSTM ネットワーク [10] | ✓ / ✓ |
| FPMC [8] | マルコフ連鎖・行列因子分解 | - / - |
| GRU4Rec [14] | Gated Recurrent Unit [14] | - / - |

提案しており、POI に関するユーザ嗜好に加えて C/I の時間的・地理的特性も併せて考慮している [2]. Feng らは Next-POI 問題だけでなく将来訪問ユーザの予測問題にも焦点を当て、POI の地理的影響を組み込む潜在表現ベースの POI2Vec を提案している [5]. 深層学習の技術を適用したシステムの研究も確認されており、Huang らは再帰型ニューラルネットワークの一種である LSTM ネットワーク [10] に基づくシステムとして ATST-LSTM を提案している [6].

C/I ログはユーザインタラクションの連続的なログの一種であるため、そのような入力に依存する Sequence-aware Recommender System (RS) [11], [12] も Next-POI 問題に適用可能である. この手法ファミリーには Session- $\{$ based/aware $\}$ RS が含まれ、代表的なシステムとして FPMC-LR のベースである FPMC [8] や、Gated Recurrent Unit (GRU) [13] を適用した深層ニューラルネットワークベースの GRU4Rec [14] などが提案されている.

本節で取り上げた Next-POI 問題を扱うことができる推薦システムの先行研究を表 1 にまとめる. 各システムが地理・時間情報を利用する場合はチェックマークで示す.

2.2 観光行動を考慮した C/I ログ活用研究

POI 推薦システムの提案研究では、LBSN から取得された C/I ログをデータセットとして評価実験が行われることが多い. その一方で、C/I ログの前処理は疎な情報の削除を目的とする処理に留まり、内包する反復性のある日常行動に関連した処理は行われていない [7]. 例えば、FPMC-LR の評価実験では Foursquare と Gowalla² の C/I ログから、C/I が 120 回未満のユーザと被 C/I 数が 5 回未満の POI を削除している [4]. Geo-Teaser と POI2Vec では C/I が 10 回未満のユーザと C/I したユーザ数が 5 人未満の POI を削除する前処理をそれぞれ実施するに留まっている [2], [5]. ATST-LSTM においても、非アクティブユーザと不人気 POI といった疎な情報のフィルタリングを目的として、C/I が 20 回未満のユーザと被 C/I 数が 10 回未満の POI を削除している [6]. 観光行動の意思決定支援を目的とした POI 推薦システムにおいて、推薦すべき観光スポットなどの POI が反復的に C/I されないことを踏まえると、反復性のある日常行動がデータセットに存在することはシステムの公平な比較に悪影響を及ぼしかねない問題である.

POI 推薦システムの評価実験で非観光行動の一種である反復性のある日常行動を考慮した研究は見当たらない一方で、C/I

ログから観光行動を抽出する方法として、ユーザ単位で抽出する 2 つの手法が提案されている. Ferreira らは Foursquare の各ユーザをユーザプロフィールを利用して観光客と居住者に分類し、両クラスのユーザ行動を分析・比較している [15]. Järv は Maeda らが提案している観光地の抽出手法 [16] を応用し、各ユーザが頻繁に C/I する重要な POI (自宅周辺・オフィスなど) をクラスタリングによって発見することで、非居住者である観光客の行動抽出に活用している [17].

上記研究は観光行動の抽出に留まり、抽出した観光行動が POI 推薦システムの性能比較に与える影響までは検証されていない. 本研究はユーザプロフィールが不要で、ツアーに含まれる POI の特徴類似性に基づくツアー単位での非観光行動、特に反復性のある日常行動の抽出法を提案する. そして、この抽出法を POI 推薦システムの評価へ応用し、非観光行動が推薦精度などの各指標に与える影響を確認する.

3 反復性のある日常行動の抽出

C/I ログはユーザと POI の集合 U, P で構成される. ここでは各ユーザの 1 日の C/I 系列を“ツアー”と呼び、ユーザ $u \in U$ はツアー集合 \mathcal{T}_u を持つ. ツアー $T \in \mathcal{T}_u$ は、 T で i 番目に C/I された POI を $p_i \in P$ として

$$T = (p_1, p_2, \dots, p_i, \dots, p_{|T|}) \quad (1)$$

のように C/I の時系列順に並べた順序組である. ツアー内で同一の POI に複数回 C/I することもあり、 T では同一 POI の複数回の出現を許容する. 全ユーザのツアー集合を \mathcal{T} とする.

3.1 行動抽出法

提案する反復性のある日常行動の抽出法は、ツアー埋め込み構築、life スコア計算、閾値設定によるラベリングの 3 つに分かれる.

3.1.1 ツアー埋め込み構築

本研究では Word2Vec (skip-gram) [9] を用いて POI の埋め込みを獲得する. 具体的には、POI・ツアーを Word2Vec における単語・文に置き換え、ある POI への C/I 前後に C/I される POI を予測するように POI 埋め込み関数 $f_\theta : P \rightarrow \mathbb{R}^d$ (d : 埋め込み次元数) のパラメータ θ を \mathcal{T} から学習する. Word2Vec のハイパーパラメータの最適値は、商品購買履歴に適用してアイテムの埋め込みを獲得している Item2Vec [18] の評価実験で採用されているジャンル (C/I ログでは POI カテゴリ) 分類タスクを用いることで決定する. 探索するハイパーパラメータとその範囲は、エポック数 $\in [1, 30]$, ベクトル次元数 $\in \{25, 50, 75, 100\}$, ウィンドウサイズ $\in \{2, 4, 6\}$, サブサンプリング閾値 $\in \{10^{-3}, 10^{-4}, 10^{-5}\}$ とする. ネガティブサンプル数は Barkan らの設定に従い 15 とする [18].

次に、ツアー T の埋め込み $t \in \mathbb{R}^d$ を、そのツアー内で C/I した POI の埋め込みの平均ベクトル

$$t = \frac{1}{|T|} \sum_{i=1}^{|T|} f_\theta(p_i) \quad (2)$$

²: <https://www.gowalla.com/>

表 2 GSCI-JP データセットの統計情報

| エリア | 主要都市 | #C/I | #user | #tour | #POI | #category |
|---------|----------|---------|--------|---------|--------|-----------|
| TOKYO | 東京・川崎・横浜 | 797,857 | 11,243 | 221,792 | 20,813 | 330 |
| OSAKA | 大阪・京都・神戸 | 241,391 | 5,744 | 64,884 | 6,916 | 274 |
| NAGOYA | 名古屋・岐阜 | 84,122 | 3,000 | 24,614 | 2,831 | 215 |
| SAPPORO | 札幌 | 23,772 | 1,039 | 7,324 | 1,039 | 152 |
| FUKUOKA | 福岡・下関 | 44,820 | 1,878 | 12,754 | 1,605 | 176 |
| SENDAI | 仙台 | 20,933 | 1,097 | 7,042 | 976 | 140 |

の計算により構築する。この計算において同じ POI への C/I が複数回ある場合は区別する。Word2Vec により構築された埋め込みは加法構成性を持ち、単語 (POI) の意味 (特徴) や単語同士の関係性の定量化を図っている。ツアー埋め込みの計算方法は、C/I した POI の特徴の平均を計算することでツアーの特徴 (コンテキスト) を定量化できる、という仮定に基づいている。

なお、自然言語処理分野では文の埋め込みを構築できるモデルとして Paragraph2Vec [19] があるが、予備実験により式 (2) で構築した埋め込みよりもツアーのコンテキストを正確に捉えられなかった。具体的には、Paragraph2Vec で構築した埋め込みによるツアー間類似度に対して、POI カテゴリ共通性や地理的類似性との適切な相関関係を確認できなかった。

3.1.2 life スコア計算

ツアー埋め込みがツアーのコンテキストを定量化できるという仮定に従うと、ツアー埋め込み間のコサイン類似度を計算することでツアー間の特徴類似性を見積もることができると考える。したがって、あるユーザの通勤・通学などの日常行動は、そのユーザが持つ他のツアーとの類似度が高く算出される性質を持つと予想する。

この性質に基づき、反復性のある日常行動を抽出するため、各ツアーの日常行動に分類することの尤もらしさを定量化する life スコアを提案する。ツアー $T \in \mathcal{T}_u$ の life スコア $l: \mathcal{T} \rightarrow [-1, 1]$ を

$$l(T) = \frac{1}{|\mathcal{T}_u| - 1} \sum_{T' \in \mathcal{T}_u \setminus \{T\}} \cos(\mathbf{t}, \mathbf{t}') \quad (3)$$

で定義する。 \mathbf{t}, \mathbf{t}' はツアー T, T' の埋め込み、 $\cos(\mathbf{x}_1, \mathbf{x}_2)$ は $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^d$ 間のコサイン類似度である。life スコアは同じユーザが持つ他ツアーとの類似度平均を計算しており、日常行動に分類することが尤もらしいツアーはスコアが高く、観光行動をはじめとする希少な行動パターンを持つツアーはスコアが低く算出されることを意図している。なお、日常行動には病院・市役所などの POI への C/I といった反復性のないツアーも想定されるが、このようなツアーのスコアは低く算出されるため、life スコアは“反復性のある”日常行動の抽出にのみ活用できる。

3.1.3 閾値設定によるラベリング

反復性のある日常行動に分類できるツアー T はユーザが日常的に C/I する POI を含むことが予想されるため、life スコア $l(T)$ は高くなると考えられる。この性質を利用し、life スコアに対する閾値 l_{th} を設定した上で、 $l(T) \leq l_{th}$ を満たすツアー T を観光行動 (ただし反復性のない日常行動も含む) trip ,

$l(T) > l_{th}$ を満たすツアー T を日常行動 life としてラベリングする。そして、 life がラベリングされたツアーを選択することで反復性のある日常行動を抽出する。

3.2 行動抽出性能の検証

提案する行動抽出法により、反復性のある日常行動をはじめとする日常行動を適切に抽出できるか確認する。抽出性能の検証には、あるツアーに対して trip と life のどちらのラベリングが適切であるか表す情報が必要となるが、そのような正解データは与えられていない。そこで、ヒューリスティックな指標として“Top-POI C/I ツアー率”と“POI ジャンル別 C/I POI 数”の 2 つの指標を提案し、抽出性能の定量化を図る。

3.2.1 検証用データセット

抽出性能の検証のため、Foursquare から収集された C/I ログのオープンデータセット Global-scale Check-in Dataset (GSCI) ³ [20], [21] を利用する。GSCI は 2012 年 4 月～2013 年 9 月に世界 77 国から収集された 3000 万件以上の C/I ログを持ち、C/I する POI には座標 (緯度・経度) やカテゴリ情報が付加されている。

本研究では、このうち日本国内のログのみ選択し、POI 座標でクラスタリングすることで各 POI を 8 つのエリアに対応付ける。さらに、C/I 数など規模の異なる 6 つのエリアを選択することでデータセットを再構成する。このデータセットを以降“GSCI-JP”と呼ぶ。GSCI-JP に対して、C/I したユニークユーザ数が 5 人未満の不人気 POI の削除、ツアー内での同一 POI への連続 C/I の集約、1 つの C/I のみ持つツアーの削除、の順で前処理を行った後の統計情報を表 2 に示す。

GSCI に含まれる POI は Foursquare が定義するカテゴリ情報を持ち、GSCI-JP においては計 352 種類のカテゴリが各 POI に付与されている。POI カテゴリは POI やツアーの特徴を理解するために有用な情報であるが、種類が多く細分化されているため解釈が複雑になりかねない。そこで、Word2Vec [9] を用いて POI カテゴリを埋め込み、この埋め込みを用いて POI カテゴリをクラスタリングし、特徴が類似する POI カテゴリを集約する。本研究ではクラスタ数を 10 として、POI カテゴリを表 3 に示すクラスタに集約し、クラスタ名をそれぞれ付与した。以後、このクラスタ名を POI カテゴリの上位概念として“POI ジャンル”と呼ぶ。

3.2.2 Top-POI C/I ツアー率

ユーザ $u \in \mathcal{U}$ が最も頻繁に C/I する POI を Top-POI と呼

3: <https://sites.google.com/site/yangdingqi/home/foursquare-dataset>

表3 クラスタリングにより決定された POI ジャンル (POI カテゴリクラス)

| POI ジャンル | POI カテゴリ例 (C/I 数トップ 3) | | |
|------------|------------------------|------------------------------|---------------------|
| RESTAURANT | Ramen / Noodle House | Coffee Shop | Japanese Restaurant |
| HOBBY | Electronics Store | Arcade | Bookstore |
| COLLEGE | University | College Academic Building | College Cafeteria |
| THEME_PARK | Theme Park | Theme Park Ride / Attraction | Resort |
| OUTDOOR | Park | Monument / Landmark | Shrine |
| TRAFFIC | Train Station | Subway | Platform |
| TRAVEL | Light Rail | Rest Area | Bike Shop |
| AIRPORT | Airport Terminal | Airport | Airport Gate |
| STORE | Mall | Department Store | Grocery Store |
| BAR | Sake Bar | Bar | Pub |

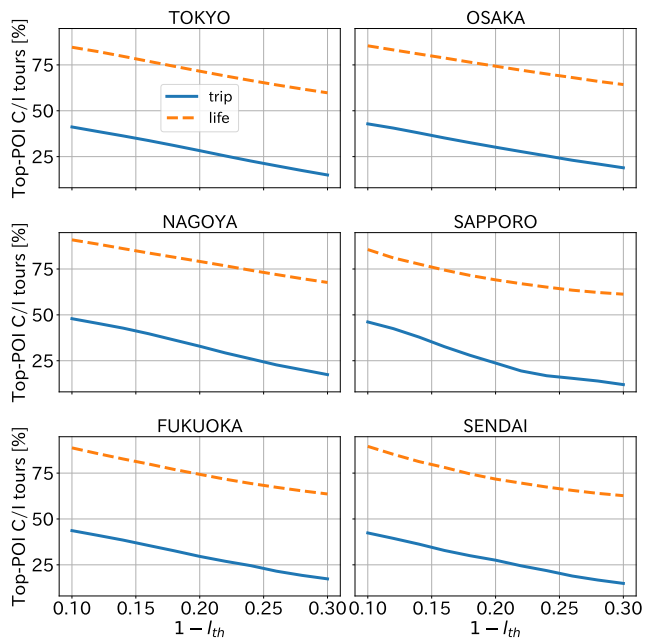


図1 Top-POI C/I ツアー率の閾値 l_{th} に対する推移

び, p_{top}^u と表記する. Top-POI C/I ツアー率はツアー集合 \mathcal{T}_u において Top-POI への C/I を含むツアーの割合を表し,

$$\frac{1}{N} \sum_{u \in \mathcal{U}} \sum_{T \in \mathcal{T}_u} \chi_T(p_{top}^u) \quad (4)$$

のように計算する (N : ツアー数). ここで χ_T は指示関数で

$$\chi_T(p_{top}^u) = \begin{cases} 1 & (p_{top}^u \in T) \\ 0 & (p_{top}^u \notin T) \end{cases} \quad (5)$$

に従い計算する. Top-POI C/I ツアー率が低い場合, 反復性のある日常行動との関連が深い学校・オフィスなどの各ユーザが頻りに訪問する POI がツアーに含まれにくいことになる.

life スコア閾値を $l_{th} \in \{0.70, 0.72, \dots, 0.90\}$ のように変化させた際の life・trip それぞれの Top-POI C/I ツアー率を図1に示す. いずれのエリアでも, life に分類されたツアーの Top-POI C/I ツアー率は trip のそれより常に 25 ポイント以上高いことが確認できる. よって, life に分類されたツアーの方が trip よりも日常行動との関連が深い POI が含まれやすく, 適切な日常行動抽出に成功していると考えられる.

3.2.3 POI ジャンル別 C/I POI 数

GSCI-JP で定義した POI ジャンル情報 (表3 参照) を利用したヒューリスティック指標として, 1 ツアーにおいて C/I された POI 数を POI ジャンル別で数え上げる, POI ジャンル別 C/I POI 数を

$$\frac{1}{N} \sum_{u \in \mathcal{U}} \sum_{T \in \mathcal{T}_u} \sum_{p \in T} \chi_h(p) \quad (6)$$

のように計算する (h : 対象ジャンル). ここで χ_h は指示関数で

$$\chi_h(p) = \begin{cases} 1 & (g(p) = h) \\ 0 & (g(p) \neq h) \end{cases} \quad (7)$$

に従い計算する. $g(p)$ は POI $p \in \mathcal{P}$ の POI ジャンルである. 同じ POI への複数回の C/I は 1 回としてカウントする.

life・trip のいずれかと関連が深い POI ジャンルの C/I POI 数は行動抽出法の性能の定量化に有用と考える. そこで, life・trip のそれぞれと関連が深い POI ジャンルとして, COLLEGE と TRAVEL を選択する. この2つの POI ジャンルについて, $l_{th} \in \{0.70, 0.72, \dots, 0.90\}$ のように変化させた際の life・trip それぞれの C/I POI 数を図2と図3に示す.

COLLEGE は日常行動との関連が深い POI ジャンルの1つである. 図2より, FUKUOKA を除くエリアの大半の区間で trip の C/I POI 数が life のそれを上回っていることが確認でき, 適切な日常行動抽出に成功していると考えられる. FUKUOKA では life と trip で C/I POI 数に大差なく, 大学施設に関連する C/I 数が全体として他エリアよりも少ないためと考える. TOKYO を除くエリアでは l_{th} を上げると life の C/I POI 数が漸増するが, $l_{th} > 0.85$ では減少していることが確認できる. これは大学での講義後に娯楽施設に行くなどして, COLLEGE の POI へ C/I するツアーが部分的に反復性のある日常行動でなくなるケースが一定数存在するためと考える.

TRAVEL は観光行動との関連が深い POI ジャンルの1つである. 図3より, 全エリアで l_{th} を下げるにつれ trip の C/I POI 数が漸増しており, 適切な日常行動抽出に成功していると考えられる. 一方, life の推移はエリアによって異なる傾向を示し, SAPPORO と FUKUOKA では $l_{th} = 0.90$ において life の C/I POI 数が trip より 0.03 以上多い. TRAVEL は車に関連する POI カテゴリを多く含むため, 日常行動での移動手段として車が使われやすい地域事情が影響していると考えられる.

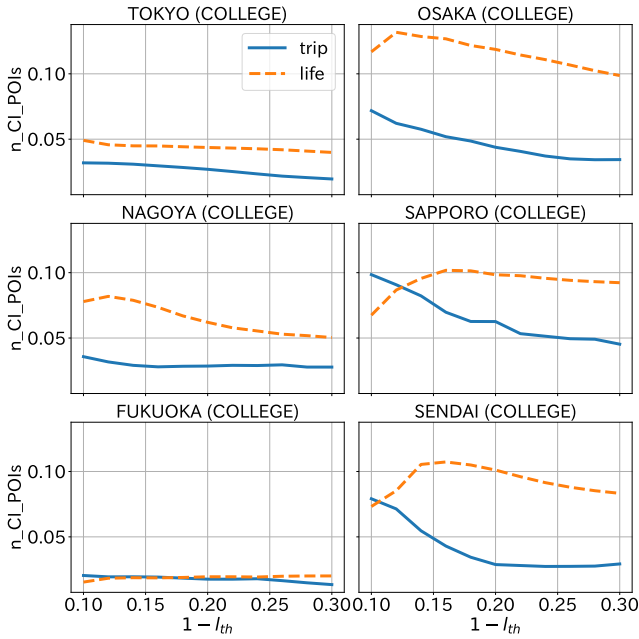


図 2 C/I POI 数 (ジャンル: COLLEGE) の閾値 l_{th} に対する推移

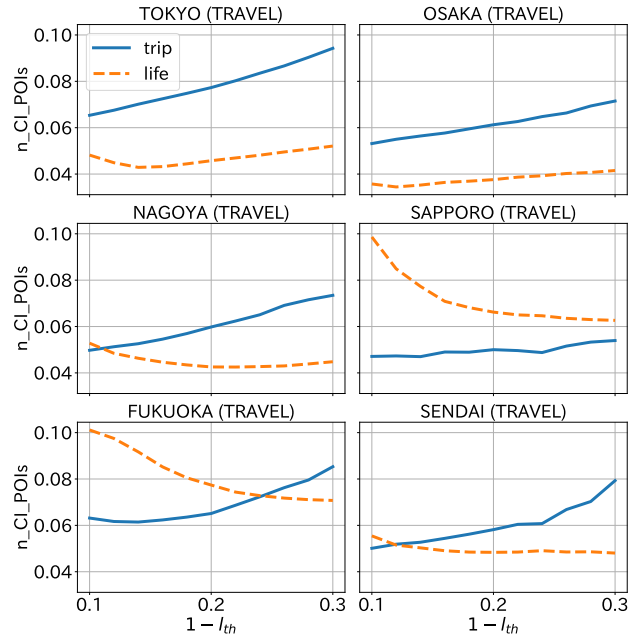


図 3 C/I POI 数 (ジャンル: TRAVEL) の閾値 l_{th} に対する推移

4 反復性のある日常行動を考慮した POI 推薦システムの評価実験

life スコア閾値 l_{th} の設定により反復性のある日常行動を抽出し、それら行動を除外した trip のツアーのみで構成される C/I ログを用いて複数の代表的な POI 推薦システムの推薦精度・多様性・新規性を評価する。これにより、反復性のある日常行動が各指標に与える影響を確認する。

4.1 POI 推薦システムの評価法

4.1.1 評価方法

ユーザーのある瞬間での次の訪問 (C/I) POI を予測する Next-POI 問題 [3] により評価する。データセットとして、3.2.1 節で示した C/I ログの GSCI-JP を使用する。C/I 時刻に基づき、訓練・検証・テストセットに分割する。分割日をシフトさせることで計 5 パターンの訓練・検証・テストセットを作成する。検証・テストセットの期間はそれぞれ 2ヶ月間とし、検証セットより前の期間を訓練セットとする。各セットの期間の詳細は表 4 に示す。検証・テストセットに含まれるツアーの先頭を除くランダムな位置の POI を 1つ選択し、ターゲット POI として予測する。

ハイパーパラメータチューニングは訓練・検証セットで行い、各パターンでの最適値を決定する。次にその最適値を用い、訓練・検証セットを連結した新しい訓練セットとテストセットを使って評価指標を計算する。各パターンで得られた評価値の平均をとることで最終的なスコアを決定する。なお、後述の POP と U-POP はハイパーパラメータを持たない。

4.1.2 評価対象システム

5つのシステムを評価対象とする。代表的な POI 推薦システムとして、表 1 から FPMC-LR [4], Geo-Teaser [2],

GRU4Rec [14] を採用する。これに加え、単純なベースラインとして、全ユーザー・各ユーザーをそれぞれ集計対象としたときの上位人気 POI を推薦する POP と U-POP を採用する。

各システムはターゲット POI を予測するため、訓練セットに含まれる POI 集合から k ヶ所の POI を選択し、それらを順位付けた推薦 POI ランキングを生成する。なお、本実験では $k = 20$ で固定する。

4.1.3 評価指標

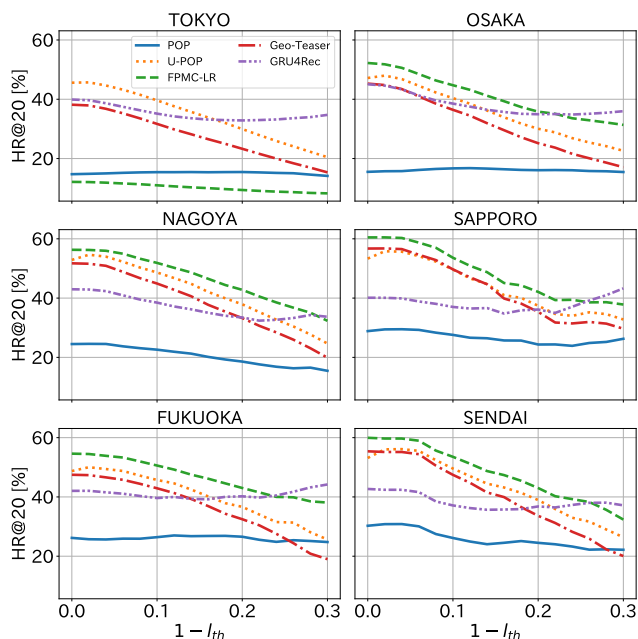
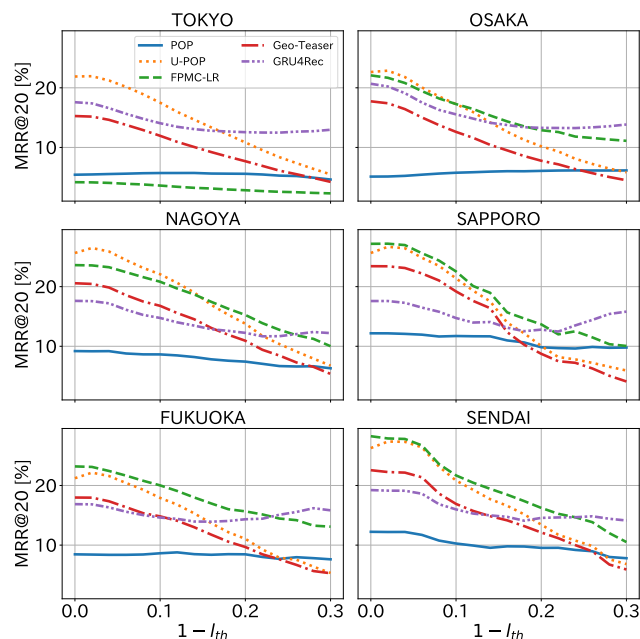
推薦精度・多様性・新規性を評価する計 4つの評価指標を採用する。推薦精度の評価には HR (Hit Ratio) @ k と MRR (Mean Reciprocal Rank) @ k を用いる。MRR は推薦 POI ランキングにおけるターゲット POI の出現順位の逆数を計算する。ハイパーパラメータの最適値の決定には MRR を基準とする。

多様性・新規性の評価には Sequence-aware RS 分野の研究で用いられている 2つの指標を採用する。多様性指標として推薦 POI ランキングのペアの非重複率を測る DIV@ k [22] を用いる。DIV が高いほど多くのターゲットに対して特定の POI が偏って推薦されにくく、バラエティに富む POI ランキングを生成できるといえる。新規性指標としてコンテキスト POI 集合と推薦 POI ランキングの非重複率を測る MCAN@ k [23] を用いる。ここでのコンテキスト POI はユーザーが過去 (訓練セットの期間) に C/I した POI を指し、MCAN はユーザーが未訪問の POI が推薦されるほどスコアが高くなる。

各評価指標は、life スコア閾値を $l_{th} \in \{0.70, 0.72, \dots, 1.00\}$ のように変化させ、各閾値で決定された trip のツアーのみを対象に集計する。 $l_{th} = 1.00$ の設定では life のラベルが付いて除外されるツアーが存在しないため、反復性のある日常行動を考慮しない場合の評価結果となる。

表 4 5 分割交差検証法を用いる訓練・検証・テストセットの期間

| Fold | Train | Validation | Test |
|------|-------------------------|-------------------------|-------------------------|
| 1 | 2012.04.04 – 2012.09.17 | 2012.09.18 – 2012.11.17 | 2012.11.18 – 2013.01.17 |
| 2 | 2012.04.04 – 2012.11.17 | 2012.11.18 – 2013.01.17 | 2013.01.18 – 2013.03.17 |
| 3 | 2012.04.04 – 2013.01.17 | 2013.01.18 – 2013.03.17 | 2013.03.18 – 2013.05.17 |
| 4 | 2012.04.04 – 2013.03.17 | 2013.03.18 – 2013.05.17 | 2013.05.18 – 2013.07.17 |
| 5 | 2012.04.04 – 2013.05.17 | 2013.05.18 – 2013.07.17 | 2013.07.18 – 2013.09.17 |

図 4 HR@20 の閾値 l_{th} に対する推移図 5 MRR@20 の閾値 l_{th} に対する推移

4.2 評価結果と考察

4.2.1 推薦精度の評価結果

各システムの life スコア閾値 l_{th} に対する HR@20 と MRR@20 の推移をそれぞれ図 4 と図 5 に示す。 $l_{th} \in \{0.70, 1.00\}$ での MRR@20 のスコアと、 $l_{th} = 1.00$ に対する $l_{th} = 0.70$ でのスコア比 r を表 5 にまとめる。表中の下線太字部はエリアと l_{th} の組み合わせにおける最高スコアを表す。

$l_{th} = 1.00$ において、U-POP と FPMC-LR は HR/MRR ともに優れる傾向にある。MRR スコアは、C/I 数が多いなど規模の大きいエリアでは U-POP、小さいエリアでは FPMC-LR が最も優れている。よって、規模の大きいエリアほど各ユーザの C/I は特定の POI に偏る傾向が強いと考えられる。GRU4Rec は TOKYO と OSAKA を除く規模の小さい 4 つのエリアにおいて、POP の次に低い HR/MRR スコアを記録している。FPMC-LR は C/I 数の最も多い TOKYO に限り、他エリアと比べて十分に機能しなかった。

life スコア閾値 l_{th} を下げていった際の推薦精度の推移を観察する。 l_{th} を下げていくことは日常行動への分類がより尤もらしいツアーが評価スコアの集計対象から除外されていくことを意味する。反復性のある日常行動は各ユーザの典型的な行動パターンであるため、ターゲット POI の予測は容易であると推測される。よって、 l_{th} を下げるほど予測が容易なターゲット POI が集計対象外になるため、推薦精度が低下すると考えられる。

実際、図 4 と図 5 より、U-POP、FPMC-LR、Geo-Teaser は l_{th} を下げることで HR/MRR ともにほとんどのケースで単調減少していることが確認できる。特に U-POP と Geo-Teaser の減少が著しく、表 5 より $l_{th} = 0.70$ では $l_{th} = 1.00$ の 25% 前後のスコアまで落ち込んでいる。おおよそ $l_{th} < 0.85$ において、FPMC-LR は U-POP や Geo-Teaser より傾きが比較的緩やかだが、それでも $l_{th} = 0.70$ でのスコアは $l_{th} = 1.00$ の 50% 前後かそれ以下まで低下している。

一方で、GRU4Rec と POP は HR/MRR ともに l_{th} を下げていった際の減少幅が比較的小さい。GRU4Rec は MRR の $l_{th} = 1.00$ において、U-POP よりも最大 8.0 ポイント低いが、 $l_{th} = 0.70$ では全エリアで他のシステムより 2.2–7.5 ポイント高いスコアを獲得している。表 5 より、SAPPORO と FUKUOKA では MRR のスコア比が $r = 0.9$ 前後を記録しており、GRU4Rec は反復性のある日常行動を除外しても推薦精度をほとんど維持することが確認できる。GRU4Rec の l_{th} の推移について、HR/MRR ともに全エリアで共通する傾向として、 $l_{th} \in [0.85, 1.00]$ では緩やかに減少する一方で $l_{th} \in [0.70, 0.85]$ では維持もしくは増加することが挙げられる。これは、他システムと同様に、GRU4Rec も life スコアが高いツアー（反復性のある日常行動）におけるターゲット POI 予測でスコアを獲得しており、life スコア分布は高い方に偏りがあり除外されるツアー数が多い影響を受けて一旦減少していると考えられる。ただ、

表 5 $l_{th} \in \{0.70, 1.00\}$ における MRR@20 スコア [%]

| エリア | TOKYO | | | OSAKA | | | NAGOYA | | | SAPPORO | | | FUKUOKA | | | SENDAI | | |
|------------|-------------|-------------|------|-------------|-------------|------|-------------|-------------|------|-------------|-------------|------|-------------|-------------|------|-------------|-------------|------|
| | l_{th} | 0.7 | 1.0 | r | 0.7 | 1.0 | r | 0.7 | 1.0 | r | 0.7 | 1.0 | r | 0.7 | 1.0 | r | 0.7 | 1.0 |
| POP | 4.62 | 5.43 | 85.2 | 6.15 | 5.11 | 120 | 6.31 | 9.20 | 68.6 | 9.79 | 12.2 | 80.4 | 7.62 | 8.46 | 90.1 | 7.80 | 12.2 | 63.8 |
| U-POP | 5.48 | 21.9 | 25.0 | 5.84 | 22.6 | 25.8 | 6.76 | 25.6 | 26.4 | 5.94 | 25.6 | 23.1 | 5.33 | 21.2 | 25.1 | 6.82 | 26.3 | 26.0 |
| FPMC-LR | 2.30 | 4.18 | 55.1 | 11.1 | 22.1 | 50.3 | 10.0 | 23.6 | 42.4 | 10.1 | 27.2 | 37.0 | 13.1 | 23.2 | 56.5 | 10.5 | 28.3 | 37.2 |
| Geo-Teaser | 4.23 | 15.3 | 27.7 | 4.50 | 17.7 | 25.4 | 5.36 | 20.6 | 26.1 | 4.10 | 23.4 | 17.5 | 5.28 | 18.0 | 29.4 | 5.91 | 22.6 | 26.2 |
| GRU4Rec | 13.0 | 17.6 | 73.7 | 13.8 | 20.7 | 67.0 | 12.2 | 17.6 | 69.3 | 15.8 | 17.6 | 89.9 | 15.8 | 16.9 | 93.9 | 14.1 | 19.2 | 73.5 |

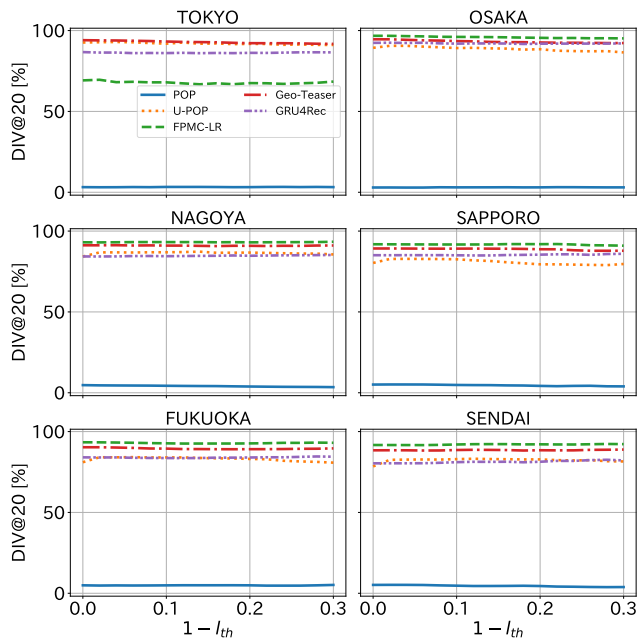


図 6 DIV@20 の閾値 l_{th} に対する推移

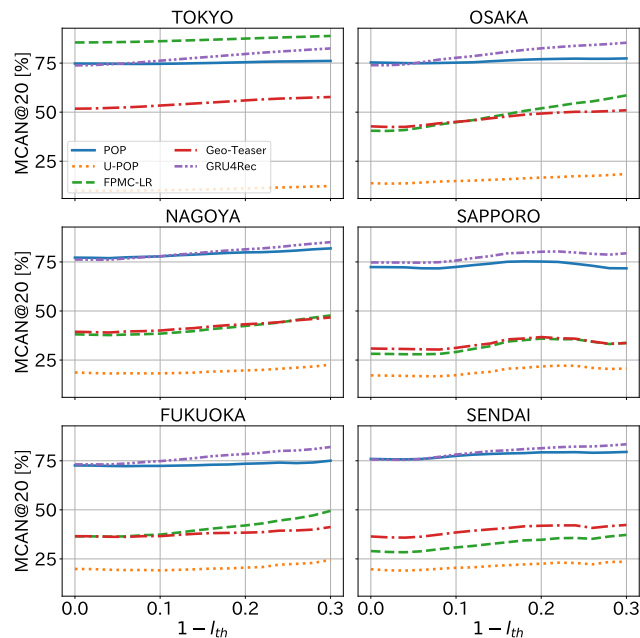


図 7 MCAN@20 の閾値 l_{th} に対する推移

life スコアが低いツアーでも精度良く推薦できるため、 l_{th} の減少に伴い除外されるツアー数が少なくなる影響を受けて次第に推薦精度を維持・増加できるようになると考える。

GRU4Rec において反復性のある日常行動を除外しても推薦精度が低下しづらい理由として、パーソナライズの程度にあると考える。GRU4Rec は Session-based RS [11] であり、推薦にはターゲット POI と同じツアーの C/I 情報のみ利用される特徴を持つ。一方、推薦精度が大きく低下した U-POP, FPMC-LR, Geo-Teaser はターゲット POI と同じユーザの C/I 情報を利用するため、パーソナライズの程度が比較的強い。したがって、パーソナライズの程度を強くすることは推薦精度の向上に貢献するものの、実際には反復性のある日常行動においてユーザが頻繁に C/I する POI の予測に成功しているに過ぎないと考えられる。POP の減少幅が小さい要因もパーソナライズの程度にあり、評価対象システムの中で最もパーソナライズの程度が低いために l_{th} の影響を受けにくいと考える。

4.2.2 多様性の評価結果

各システムの life スコア閾値 l_{th} に対する DIV@20 の推移を図 6 に示す。反復性のある日常行動を除外することによる多様性指標への影響はほとんど確認できない。反復性のある日常行動においてはターゲット POI が重複することもあり得るため、推薦 POI ランキングの重複率が高くなり、多様性が低下する

と予想する。しかし、DIV はユーザを区別せずに推薦 POI ランキングのペアの非重複率を測るため、スコア上ではその傾向は表れていない。

U-POP は OSAKA, SAPPORO, FUKUOKA で、 l_{th} の減少に伴い他システムよりもスコアが数ポイント低下していることが観察できる。これは、U-POP は同じユーザであれば同じ推薦 POI ランキングを出力するシステムであり、先の 3 エリアでは l_{th} の通減によりユーザ数が減少しやすく、非重複率が 0 になる同じユーザのペアの割合が増加したからであると考えられる。

4.2.3 新規性の評価結果

各システムの life スコア閾値 l_{th} に対する MCAN@20 の推移を図 7 に示す。新規性はパーソナライズの程度との関連が深く、 l_{th} の設定にかかわらず GRU4Rec や POP といったパーソナライズが弱いシステムのスコアが高いことが確認できる。

life スコア閾値 l_{th} を下げていくと、FPMC-LR, Geo-Teaser, GRU4Rec は数ポイントから 10 ポイント程度増加している。このことは、反復性のある日常行動では日常的な行動エリア内の POI が推薦されやすく、そのような POI は既に C/I している可能性が高いため l_{th} を下げるとスコアが高くなると考える。

4.2.4 評価実験のまとめ

life スコア閾値 l_{th} を下げていくことで反復性のある日常行動を一定数除外した際、多くのシステムは推薦精度が低下して

いくが、GRU4Rec はエリアにかかわらず除外前の精度を比較的高い水準で維持することが明らかになった。 $l_{th} = 1.00$ から $l_{th} = 0.70$ への推移により、推薦精度の観点でシステムの相対的な良し悪しが変化しており、反復性のある日常行動がシステムの性能比較に影響を与えることが確認できる。一方、 l_{th} はシステムの多様性と新規性の比較に目立った影響を与えない。ただ、新規性は l_{th} を下げていくことで全体的にスコアが増加しており、反復性のある日常行動の存在が新規性を過小評価する要因になると推測される。

5 おわりに

本研究では、反復性のある日常行動を除外した C/I ログを用いて複数の代表的な POI 推薦システムを評価し、推薦精度・多様性・新規性に与える影響を確認している。反復性のある日常行動を除外するため、ツアーに含まれる POI の特徴類似性を考慮し、そのような行動をツアー単位で抽出する方法を提案している。提案する抽出法はヒューリスティックな指標によって日常行動を適切に抽出できることを確認している。POI 推薦システムの評価結果として、反復性のある日常行動を一定数除外した際、多くのシステムと異なり、GRU4Rec は推薦精度を比較的高い水準で維持しており、反復性のある日常行動がシステムの性能比較に影響を与えることを確認している。一方、多様性・新規性に対する反復性のある日常行動の顕著な影響は確認されていない。

今後の方向性として、病院に行くなど観光行動ではないが反復性のない日常行動も考慮できるような行動抽出法の改善、適切なヒューリスティック指標を追加するといった抽出性能の検証方法の改善、より幅広いアプローチのシステムを評価対象とすることにより観光行動のモデリングに適した POI システムの特性の体系化を試みることなど検討している。

謝 辞

本研究は JSPS 科研費 JP21H03553 の助成を受けたものです。

文 献

- [1] Japan Tourism Agency, “White paper on tourism in Japan, 2021 (Summary),” Japanese Ministry of Land, Infrastructure, Transport and Tourism, 2021.
- [2] S. Zhao, T. Zhao, I. King, M. R. Lyu, “Geo-Teaser: Geotemporal sequential embedding rank for Point-of-interest recommendation,” Proc. of the 26th Int. Conf. on World Wide Web Companion, pp. 153–162, 2017.
- [3] H. Werneck, N. Silva, M. Viana, A. C. Pereira, F. Mourão, L. Rocha, “Points of interest recommendations: Methods, evaluation, and future directions,” Inf. Syst., Vol. 101:101789, 2021.
- [4] C. Cheng, H. Yang, M. R. Lyu, I. King, “Where you like to go next: Successive Point-of-interest recommendation,” Proc. of the 23rd Int. Jt. Conf. on Artif. Intell., pp. 2605–2611, 2013.
- [5] S. Feng, G. Cong, B. An, Y. M. Chee, “POI2Vec: Geographical latent representation for predicting future visitors,” Proc. of the AAAI Conf. on Artif. Intell., Vol. 31,

- No. 1, 2017.
- [6] L. Huang, Y. Ma, S. Wang, Y. Liu, “An attention-based spatiotemporal LSTM network for next POI recommendation,” IEEE Trans. on Serv. Comput., Vol. 14, Issue. 6, pp. 1585–1597, 2019.
- [7] 榎木 佑真, 岡本 一志, “非観光行動の影響を考慮した POI 推薦システムの一評価法,” Web インテリジェンスとインタラクショナル研究会 予稿集, Vol. 18, No. 7, pp. 31–34, 2022.
- [8] S. Rendle, C. Freudenthaler, L. Schmidt-Thieme, “Factorizing personalized Markov chains for next-basket Recommendation,” Proc. of the 19th Int. Conf. on World Wide Web, pp. 811–820, 2010.
- [9] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, J. Dean, “Distributed representations of words and phrases and their compositionality,” Proc. of the 26th Int. Conf. on Neural Inf. Process. Syst., Vol. 2, pp. 3111–3119, 2013.
- [10] S. Hochreiter, J. Schmidhuber, “Long short-term memory,” Neural Comput., Vol. 9, No. 8, pp. 1735–1780, 1997.
- [11] S. Latifi, N. Mauro, D. Jannach, “Session-aware recommendation: A surprising quest for the state-of-the-art,” Inf. Sci., Vol. 573, pp. 291–315, 2021.
- [12] M. Quadrana, P. Cremonesi, D. Jannach, “Sequence-Aware Recommender Systems,” ACM Comput. Surv., Vol. 51, Issue. 4, No. 66, pp. 1–36, 2018.
- [13] K. Cho, B. Van Merriënboer, D. Bahdanau, Y. Bengio, “On the properties of neural machine translation: Encoder-decoder approaches,” arXiv, arXiv:1409.1259, 2014.
- [14] B. Hidasi, A. Karatzoglou, L. Baltrunas, D. Tikk, “Session-based recommendations with recurrent neural networks,” Int. Conf. on Learn. Represent., 2016.
- [15] A. P. G. Ferreira, T. H. Silva, A. A. F. Loureiro, “Beyond sights: Large scale study of tourists’ behavior using Foursquare data,” IEEE Int. Conf. on Data Min. Workshop, pp. 1117–1124, 2015.
- [16] T. N. Maeda, M. Yoshida, F. Toriumi, H. Ohashi, “Extraction of tourist destinations and comparative analysis of preferences between foreign tourists and domestic tourists on the basis of geotagged social media data,” ISPRS Int. J. Geo-Inf., Vol. 7, Issue. 3, 99, 2018.
- [17] P. Järv, “Mining behavior from Foursquare check-ins,” Tallinn University of Technology, 2019.
- [18] O. Barkan, N. Koenigstein, “Item2Vec: Neural item embedding for collaborative filtering,” IEEE 26th Int. Workshop on Mach. Learn. for Signal Process., pp. 1–6, 2016.
- [19] Q. Le, T. Mikolov, “Distributed representations of sentences and documents,” Int. Conf. on Mach. Learn., Vol. 32, pp. 1188–1196, 2014.
- [20] D. Yang, D. Zhang, B. Qu, “Participatory cultural mapping based on collective behavior data in location-based social networks,” ACM Trans. on Intell. Syst. and Technol., Vol. 7, Issue. 3, pp. 1–23, 2016.
- [21] D. Yang, D. Zhang, L. Chen, B. Qu, “NationTelescope: Monitoring and visualizing large-scale collective behavior in LBSNs,” J. of Netw. and Comput. Appl., Vol. 55, pp. 170–180, 2015.
- [22] L. Hu, L. Cao, S. Wang, G. Xu, J. Cao, Z. Gu, “Diversifying personalized recommendation with user-session context,” Proc. of the 26th Int. Jt. Conf. on Artif. Intell., pp. 1858–1864, 2017.
- [23] S. Wang, L. Hu, L. Cao, X. Huang, D. Lian, W. Liu, “Attention-based transactional context embedding for next-item recommendation,” Proc. of the AAAI Conf. on Artif. Intell., Vol. 32, No. 1, pp. 2532–2539, 2018.