

交通案内サイト利用履歴を用いたユーザの移動予測

丸山 直樹* 若宮 翔子* 細川 侑士* 河合 由起子* 秋山 豊和* 角谷 和俊†
 Naoki Maruyama Shoko Wakamiya Yuji Hosokawa Yukiko Kawai Toyokazu Akiyama Kazutoshi Sumiya

1. まえがき

Twitter や Facebook のようなソーシャルネットワークに投稿されている大量のユーザ発信コンテンツを分析し、有用な情報を抽出する研究 [1] が盛んに行なわれている。特に、災害のようなイベント発生時に、被災支援のために活用するサービス [2] が注目を集めている。しかし、このようなサービスは大量のユーザに同一の情報を配信しており、各ユーザに適した情報を迅速に提供するまでには至っていない。各ユーザに適切な情報を提示するためには、ユーザが将来どこにいるかという位置推定を行う必要がある。これに対し、我々の先行研究 [3] では、ユーザに対して適切な情報を素早く提供することを目的として、日本国内の主要交通案内サイトにおけるユーザの操作・閲覧履歴の URI を分析し、災害等のイベントが発生した場所にいるユーザが居住者か訪問者かを判定し、居住ユーザと訪問ユーザごとに適した避難情報を提示するシステムを開発してきた。しかしながら、交通案内サイトから抽出される地名と時刻の情報から、将来の各ユーザの行動を推定することができず、個々のユーザに適した情報を提示するまでには至っていなかった。そこで、本研究では、ユーザの交通案内サイト利用履歴を分析し、ユーザの移動を確率的に推定する手法を提案する。提案する移動推定手法は、交通案内サイトの利用履歴を用いて「存在確率」と「行動確率」を抽出する。存在確率の抽出は、地名と時刻の情報を用いて、最尤法により各時刻における移動時間長を重みとして、各地名の尤度を算出する。また、既知情報に対して操作・閲覧が行なわれない場合、ユーザの嗜好性および習慣性に基づき欠損データを補間し、行動確率を算出する。欠損データ補間手法は、1) ユーザの過去の操作・閲覧履歴の類似性を用いた手法と 2) 群衆行動との類似性を用いた手法を適用する。そして、これらの手法で算出した存在確率と行動確率から、将来にわたるユーザの移動を予測する。本論文では、将来におけるユーザの存在確率および行動確率を算出する手法について述べ、実装したシステムの有用性を検証する。

本論文の構成は以下の通りである。2章ではシステムの概要を紹介する。3章において、交通案内サイト利用履歴から地名を抽出するための手法を紹介し、4章では、ユーザの移動を推定する具体的な手法を述べる。5章において、被験者を用いて実験を行った結果について述べ、6章で関連研究を紹介する。最後に、7章でまとめと今後の課題について述べる。

2. システム概要

本システムでは、交通案内サイトに対するユーザの利用履歴情報を用いてユーザの移動を推定する。移動

を予測する事で、イベントが発生した際に、各ユーザに適したイベント情報を Web ブラウザ上にポップアップ提示可能となる。例えば、図 1 のユーザ A のように災害発生時に災害発生場所である X 市にいるユーザに対して避難警告情報を提示するだけでなく、図 1 のユーザ B のように現在 X 市にはいないが、将来 X 市を訪問する可能性が高いユーザに対しても警報情報を提示し、注意を促すことができる。

提案システムの概要を図 1 に示す。まず、ユーザは Firefox の拡張機能をインストールする。その際、ユーザは居住地である地域名を選択し、登録する。この拡張機能により、ユーザのブラウジングの閲覧履歴情報である URI を取得する。一般的に、ブラウザを用いた意思決定において、ユーザは交通案内サイト等の「フォーム」を用いて往路探索の決定を行うことが多いため、全ページを解析せず query を含む URI とその入出力 Web ページのみを解析対象とする。これにより、コストの削減とユーザのプライバシーの考慮につながる。次に、取得した URI の query 部分に交通情報が含まれているかを判定し、出発地、到着地、日付、時刻の情報を抽出する。query 部分のみでこれらの情報を抽出できない場合は、query を入力した Web ページおよび出力結果の Web ページを分析して抽出する。本手法では、抽出されたこれら地名と時刻の情報を用いて、最尤法により現在地の時刻における各地名に対する存在確率を算出する。そして、ユーザの嗜好性および習慣性により既知である情報に対して操作・閲覧が省略される場合、欠損データ補間手法を用いて行動確率を算出する。これら算出した存在確率および行動確率に基づき、将来におけるユーザの移動を予測する。最後に、災害等のイベントが発生した際、ユーザに適した情報を提示する。

3. Web 閲覧履歴を用いた地名抽出

本研究では、交通案内サイトにおける交通情報や宿泊情報の閲覧履歴からユーザの行動を推定する。一般的に、ユーザはブラウザを用いて意思決定を行う際に交通案内サイトや宿泊サイトの「フォーム」を用いて往路探索や宿泊場所の決定を行う。本論文では、ユーザがフォームに入力した後に得られる URI そのものから出発地、到着地、到着時刻等の情報を抽出する。ユーザ行動推定は URI のうちフォームに入力された query 部分を用いる。query は name 値=value 値のペアの連続値で構成されている。解析対象となる URI は query を含む URI とし、そのフォーム入力ページ、出力ページのみを対象とする。具体的な抽出手順を以下に示す。

1. URI の query 部分の value 値を形態素解析する。
2. 形態素解析結果の品詞が「名詞:地名」のキーワードを抽出し、value 値を地名と判定する。

*京都産業大学 コンピュータ理工学部

†兵庫県立大学 環境人間学部

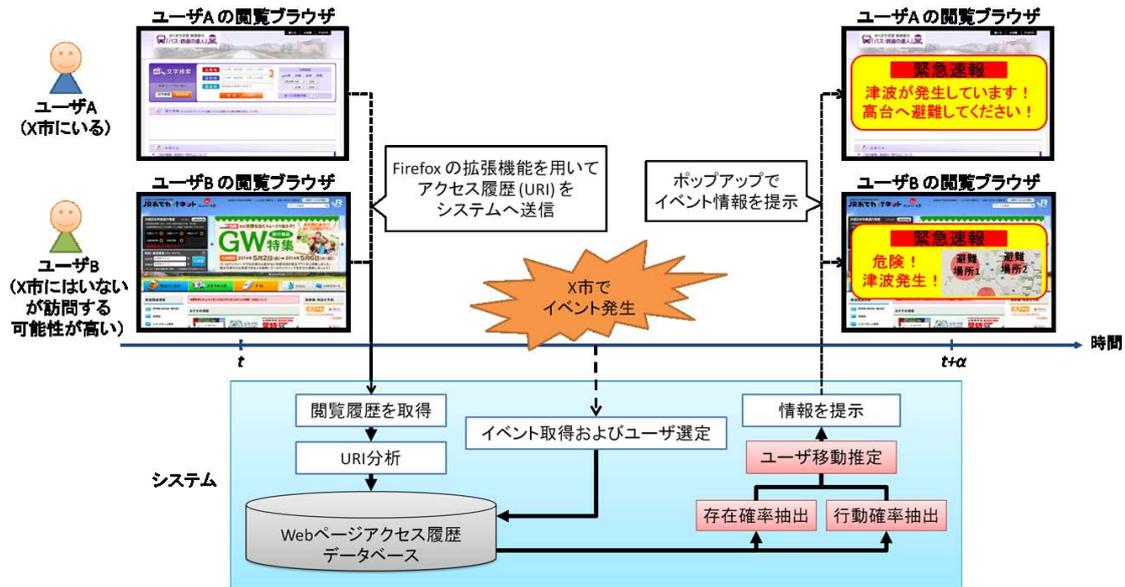


図 1: システム概要図

3. 取得した地名の name 値がフォーム入力ページの HTML ソース内の name 値と一致し、かつ同じ行に「出発/乗車」あるいは「到着/目的/下車」というキーワードを含んでいた場合、取得した地名が「出発地」あるいは「到着地」と判定する。
4. 同じ行に「出発/乗車」あるいは「到着/目的/下車」というキーワードを含まない場合、取得した2つの地名の value 値に対する name 値がフォーム入力ページの HTML ソース内の name 値と一致する行の間に「出発/乗車」あるいは「到着/目的/下車」というキーワードを含むか判定する。
5. 「出発/乗車」というキーワードを含む場合、HTML ソース内の2つの name 値を含む行間と同じ行数を name 値を含む行の前後で「出発/乗車」あるいは「到着/目的/下車」というキーワードが存在するか判定する。
6. HTML ソース内の1つ目の name 値を含む行より前で「出発/乗車」、「到着/目的/下車」というキーワードを含まず、かつ HTML ソース内の2つ目の name 値を含む行より後に「到着/目的/下車」を含む行が存在すれば1つ目の name 値に対する value 値を「出発地」、2つ目の name 値に対する value 値を「到着地」と判定する。
7. URI の query 部分の value 値に「/」あるいは「:」を含む場合は「日時」あるいは「時刻」と判定する。また、出力ページの HTML ソース内に「selected」で選択されたキーワードを抽出し、キーワードに「年」、「月」、「日」、「時」、「分」というキーワードを含んでいる場合、「日時」あるいは「時刻」と判定する。

表 1: ユーザの入力情報

| ユーザ ID | 年 | 月 | 日 | 時 | 出発地 | 到着地 |
|--------|------|---|----|----|-----|-----|
| A | 2014 | 3 | 10 | 12 | 追浜 | 池袋 |
| A | 2014 | 3 | 10 | 12 | 追浜 | 宇都宮 |
| A | 2014 | 3 | 10 | 12 | 追浜 | 霞ヶ関 |
| B | 2014 | 3 | 10 | 14 | 板橋 | 堺 |
| B | 2014 | 3 | 10 | 14 | 板橋 | 高の原 |
| C | 2014 | 3 | 10 | 14 | 宇都宮 | 霞ヶ関 |

以上よりユーザの入力、閲覧・操作履歴から抽出されるデータ例を表1に示す。

4. ユーザの移動推定

実空間におけるユーザの移動を推定するために、交通案内サイトの利用履歴から抽出されたユーザの入力情報である日時、出発地名、到着地名を用いて存在確率と行動確率を算出する。

4.1. 最尤法による入力情報に基づいた存在確率

3で得られた情報から抽出した地名がどれだけの頻度でURIに出現したかを、その地への移動の尤もらしさとして算出する。総ページ閲覧回数を n 、その地名がURIに出現した回数を k として以下の対数尤度関数に代入し、 $L(p) = 0$ となるような p の値を算出する。

$$L(p) = \log(p C_k) + k \log p + (n - k) \log(1 - p) \quad (1)$$

また、同時刻で閾値以上の地名が複数抽出された場合は、前後の地名との移動時間を検出し、移動時間が長ければ存在確率が低くなる。例えば、 T 時に到着地として東京、大阪という2つの地名が抽出されており、 $T-X$

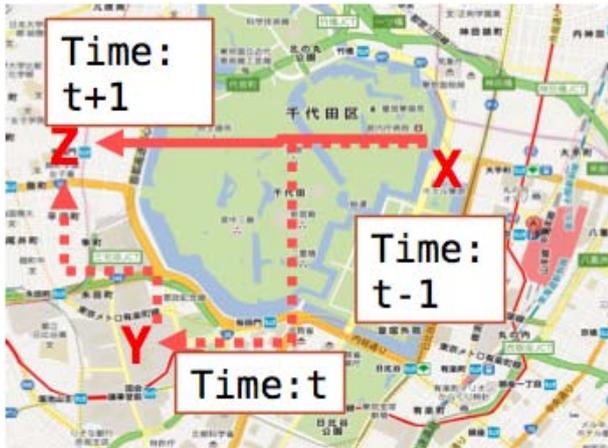


図 2: 行動推定図

時間前の到着地として奈良が算出されているとする。この場合、奈良から大阪の移動時間は X 時間よりも短い、奈良から東京への移動時間は X 時間よりも長いとすると、時刻 T のユーザの存在確率は、大阪の方が高くなる。同様に、 $T+Y$ 時の到着地に対しての移動時間も時刻 T の存在確率を算出するための重みとして用いる。

4.2. 欠損データ補間手法に基づく行動確率

前節の存在確率に加えて、操作・閲覧履歴が欠損した場合に各ユーザの過去の履歴（習慣性）に基づき、欠損データを補間する。このような欠損データ具体例として、図 2 の X 地点から Z 地点への移動情報が得られるとする。 Y 地点は今回は入力されていないが、ユーザの過去の移動履歴として、 X 地点から Z 地点への移動の際に Y 地点を通ることが分かっている場合、 Y 地点への経路を 1 で述べた 1) で補間する欠損データとなる。また、2) で補間する欠損データは、 X 地点から Z 地点へ移動する多くのユーザが立ち寄るような Y 地点のデータであり、ユーザが入力しなくても Y 地点へ訪れる可能性があるとして判断される。提案手法では、ユーザの過去の操作・閲覧履歴の類似性を検出し、類似性の高い地名集合のうち、欠損している地名への移動確率を算出する。具体的には、過去に $\{X, Y, Z\}$ の地名入力を行った後に $\{X, Z\}$ の地名を入力した場合、 $\{X, Y, Z\}$ の行動確率集合を $A = \{pastPX, pastPY, pastPZ\}$ 、その後の入力 $\{X, Z\}$ の行動確率集合を $B = \{nowPX, nowPY, nowPZ\}$ とし、式 (2) により相関係数を算出する。なお、 A の要素は 2 で抽出した存在確率、 B_{ave} は存在確率の平均値、入力があった $nowPY$ は 0 とする。

$$\frac{\sum_{i=1}^n (A_i - A_{ave})(B_i - B_{ave})}{\sqrt{\sum_{i=1}^n (A_i - A_{ave})^2} \sqrt{\sum_{i=1}^n (B_i - B_{ave})^2}} \quad (2)$$

これに $pastPY$ 重み付き和として足し合わせることで、今回のデータで欠損している地名 Y への行動確率 $nowPY$ が算出される。ただし、何度も同じ地名を調べていると判定された場合、はじめから $nowPY$ を 0 とはせず、

今までに調べられた回数に基づき 0 より大きい値を与えることで、よりユーザの嗜好性に近い欠損データの補間を行なうことが可能であると考えられる。同様に、嗜好性が類似するユーザの特徴に基づき欠損データを補間する。ここでは、他ユーザとの類似性を相関係数によって算出する。これにより、ユーザが入力していない地名に対して、嗜好性が類似しているユーザが移動した地名集合のうち欠損している地名への移動確率が算出可能となる。

5. 実験

開発したシステムの有効性を検証するため、以下の実験を行った。

1. ユーザ操作履歴に基づいた存在確率の妥当性の検証
2. 欠損データに対する補間手法の妥当性の検証

事前実験として、国内の抽出対象となる交通案内サイト 8 件に対する日時と出発地目的地抽出精度を検証した。これらのサイトより、出発地、到着地、日付、時刻の 4 項目の抽出を行った結果、ハイパーダイヤを除く 7 件のサイトで抽出が成功した。これらの情報から各時刻におけるユーザの行動推定を行う。

表 2: 実験に使用した交通案内サイトのリスト

| | 交通案内サイト |
|---|-------------|
| 1 | ジョルダン |
| 2 | Yahoo!路線情報 |
| 3 | 駅探 |
| 4 | goo 路線 |
| 5 | Google 乗換案内 |
| 6 | エキサイト乗り換え案内 |
| 7 | ハイパーダイヤ |
| 8 | NAVITIME |

5.1. 実験環境

Web サイトを操作・閲覧するブラウザは JavaScript で記述した拡張機能をインストールした Firefox バージョン 18.0 を利用し、ユーザの行動推定に必要なアクセス履歴を取得する。データベースは SQL version 3.6.12 を用いた。取得した URI は WebSocket 通信を採用し、リアルタイムに取得・管理する。被験者は関西在住の 20 代から 50 代の男女 8 名である。

5.2. ユーザ操作履歴に基づいた存在確率の妥当性の検証

実験対象者には、出発地が奈良駅であることを告げ、候補地を提示する (表 3)。被験者は、出発地から候補地までのアクセスを表 2 のサイトを用いて自由に調べる。調べた結果から、被験者は目的地を選択する。ただし、1 回目は候補地が 1 カ所なので、目的地も同じである。本実験では、複数の選択肢の中から 1 つを選択する行為に対する実験を想定しており、2 回目以降は、各候補地の距離は同一県内とした。提案する存在確率はこれらの閲覧履歴を用いて算出されるが、本実験で

表 3: 提示情報 1

| 回数 (回目) | 候補地 | | | | |
|------------|---------|---------|----|----|---|
| | 1 回目 | 福知山 | - | - | - |
| 2 回目 | 池田 (大阪) | 守口 (大阪) | - | - | - |
| 3 回目 | 西宮 | 姫路 | 明石 | - | - |
| 4 回目 | 近江八幡 | 草津 | 彦根 | 米原 | - |

表 4: ユーザ A (49 歳) の実験結果

| 回数 (回目) | PV (PV) | 操作時間 (分) | PV 割合 (秒/ページ) | 目的地 | 存在 確率 | 判定 |
|------------|------------|-------------|------------------|-----|----------|----|
| 1 | 8 | 2 | 25 | 福知山 | 1.00 | ○ |
| 2 | 14 | 3 | 21 | 守口 | 0.67 | ○ |
| 3 | 24 | 5 | 21 | 姫路 | 0.50 | ○ |
| 4 | 33 | 8 | 24 | 彦根 | 0.45 | × |

は式 (1) を用いて算出された結果のうち閾値を 0.5 以上を目的地と判定した。各ユーザの試行結果を表 4 から表 7 に示す。正解率は全体の 81% であった。候補地が 2 つ以内の場合においては 100% 正解となった。

5.3. 欠損データに対する補間手法の妥当性の検証

次に、他ユーザの抽出データとの相関性を用いたデータ補間手法により、他の全てのユーザから重み付き和として各候補地に対する移動確率を算出し、それを 5.2 節での値に付与することで、欠損データに対する補間手法の妥当性を検証する。なお、本手法は既知のデータに対するユーザの振る舞いから欠損データを補間するため、5.2 節の被験者の居住権内を対象とした実験結果に対して検証する。

表 8 に、表 4 から 7 に示した結果のうち、不正解であったデータ 3 つに着目し 5.3 節のデータ補間手法を用いて算出した移動確率結果を示す。結果より、5.2 節だけでは不正解であったデータのうち、2 つを正しく、正解データと判定できた。

5.4. 考察

実験結果から、候補地が少ない場合においては高い精度で行動の推定に成功しているが、候補地が増えると精度が低くなるのが分かる。これは、候補地が多くなるとユーザの PV が増加し、目的地となる地名以外を調べたページでの操作回数が増え、目的地の推定が難しくなることが考えられる。例えば、交通手段が複数ある候補地では、各手段での移動について一通り調べた上で、値段や乗り換え回数などを比較しなければならぬため、PV が増加し、その候補地での存在確率が高くなりやすい。また、その候補地が被験者にとって身近であるかどうか PV に影響を与える可能性がある。身近であるほど詳しい情報が必要なくなるため PV が少なくなる事が考えられる。そのため、各ユーザの操作履歴だけでは、信憑性の高いデータを算出することが難しいことがわかる。そこで、他ユーザの抽出データを用いることでより信憑性の高いデータを算出することができるが 5.3 節の結果から考察される。また、結果の中の 3 つの不正解データのうちの 1 つは 5.3 節の手法を用いても移動確率が低いままであった。

表 5: ユーザ B (52 歳) の結果

| 回数 (回目) | PV (PV) | 操作時間 (分) | PV 割合 (秒/ページ) | 目的地 | 存在 確率 | 判定 |
|------------|------------|-------------|------------------|-----|----------|----|
| 1 | 11 | 3 | 27 | 福知山 | 1.00 | ○ |
| 2 | 21 | 4 | 19 | 池田 | 0.73 | ○ |
| 3 | 26 | 5 | 19 | 西宮 | 0.57 | ○ |
| 4 | 25 | 5 | 20 | 草津 | 0.50 | ○ |

表 6: ユーザ C (26 歳) の結果

| 回数 (回目) | PV (PV) | 操作時間 (分) | PV 割合 (秒/ページ) | 目的地 | 存在 確率 | 判定 |
|------------|------------|-------------|------------------|-----|----------|----|
| 1 | 4 | 1 | 25 | 福知山 | 1.00 | ○ |
| 2 | 12 | 3 | 25 | 守口 | 0.50 | ○ |
| 3 | 24 | 5 | 21 | 西宮 | 0.38 | × |
| 4 | 24 | 5 | 21 | 彦根 | 0.54 | ○ |

表 7: ユーザ D (22 歳) の結果

| 回数 (回目) | PV (PV) | 操作時間 (分) | PV 割合 (秒/ページ) | 目的地 | 存在 確率 | 判定 |
|------------|------------|-------------|------------------|-----|----------|----|
| 1 | 11 | 1 | 10 | 福知山 | 1.00 | ○ |
| 2 | 22 | 4 | 18 | 守口 | 0.50 | ○ |
| 3 | 15 | 3 | 20 | 明石 | 0.75 | ○ |
| 4 | 18 | 3 | 17 | 米原 | 0.33 | × |

表 8: 5.3 節の検証結果

| ユーザ ID | 目的地 | 5.2 節での 移動確率 | 5.3 節での 移動確率 | 判定 |
|--------|-----|-----------------|-----------------|----|
| A | 彦根 | 0.454 | 0.567 | ○ |
| C | 西宮 | 0.375 | 0.5125 | ○ |
| D | 米原 | 0.330 | 0.380 | × |

そこで表 8 の不正解データの目的地に着目すると、5.3 節を用いて閾値をこえたデータの目的地は、表 4 から 7 の他ユーザの目的地として選ばれていることが見て取れる。しかし、ユーザ D の目的地として選んだ米原は他のユーザが目的地として選ばなかった地名であることがわかる。この結果により、この手法では、類似した嗜好性を持つユーザが他のユーザの中に含まれていない場合において、データ補間の精度に影響が出てくることが考えられる。

5.5. 未知の目的地に対する存在確率算出の検証

5.4 節の考察をふまえ、データ補間による移動推定精度の向上を計るため、被験者を増やした追加実験を行った。また、前項までの実験では被験者の在住地域である近畿圏内を候補地とし同一県内のみを候補地としていたが、追加実験では被験者の在住地域外も候補地とし、より広い移動範囲の実験を行い前項の実験結果と比較する。被験者への提示情報は表 9 とし、その他の条件は前項までの実験と同様である。5.2 節を用いた正解率は全体の 75% であった。前項の実験より精度が低い原因として、5.4 節で述べた被験者の候補地への身近さが影響していると考えられる。被験者の在住地域外の候補地が出現した場合や移動の範囲が広い場合、様々な手段での移動を選択してみる必要があるため、目的地として選定する地名以外のページ閲覧回数が増えてしまう傾向にあると考えられる。

表 9: 提示情報 2

| 回数 | 候補地 | | | |
|------|------|------|-----|----|
| 1 回目 | 天王寺 | - | - | - |
| 2 回目 | 嵐山 | 大和郡山 | - | - |
| 3 回目 | 長野 | 倉吉 | 松山 | - |
| 4 回目 | 豊前長洲 | 佐賀 | 佐世保 | 人吉 |

表 10: 5.2 節を用いた正誤表 1

| 回数 | ユーザ A | ユーザ B | ユーザ C | ユーザ D |
|------|-------|-------|-------|-------|
| 1 回目 | ○ | ○ | ○ | ○ |
| 2 回目 | ○ | ○ | ○ | ○ |
| 3 回目 | ○ | × | × | ○ |
| 4 回目 | ○ | × | × | × |

そして、5.3 節のデータ補間による移動確率の算出では、正解率は全体の 96.8% となった。被験者を増やしたことにより、データ補間の精度向上が見込めた。今回の実験では過去のデータを用いた実験は行っていないが、今後はこれらのデータを過去の履歴として管理し、同ユーザの移動確率への付与を行い、嗜好性の類似する他ユーザへのデータ補間にも活用する必要がある。さらに、異なる時刻に対して複数の目的地が出現した場合のユーザの位置推移を推定する必要がある。

6. 関連研究

Twitter の公共アカウントのツイート内容とフォローワーの推移を分析し、災害時における Twitter での公共機関のアカウントのガイドラインの作成を行った研究 [4] や Twitter に投稿された実生活情報から有用性の高いものを抽出し局面に応じた記事をユーザに提示するシステム [5] が提案されている。これらが災害等のイベントが発生した際に多くのユーザに同一の情報を提示するのに対し、本研究ではユーザの位置推定に基づき各ユーザに適した情報を提示する点で異なる。また、Twitter におけるリツイートをベースとしたユーザ間の情報伝播の関連性について分析し、ユーザが関心をもち、より緊急性の高い記事の情報伝播について分析した上で、災害時における Twitter でのリアルタイム情報発信の可能性について考察したもの [6] もある。アクセス履歴からユーザの行動を推定する手法に、Web サイトのアクセス履歴からユーザの傾向を読み取り、次にどのページを閲覧するかを予測するシステム [7] や機械学習手法により Web 上におけるユーザの行動推定を行うシステム [8] 等がある。これらが次に閲覧するであろうページの予測という仮想的な行動の推定を目的としているのに対し、本研究ではユーザの位置推定という現実の行動の予測を行い、それに基づく適切な情報をユーザに提示するという点を目的としている点で異なる。GPS などの位置探知デバイスから取得された情報を用いてユーザ行動を推定する研究 [9][10] も行われている。これに対し、我々の提案手法では過去に訪れていない場所への移動を推定することができ

表 11: 5.2 節を用いた正誤表 2

| 回数 | ユーザ E | ユーザ F | ユーザ G | ユーザ H |
|------|-------|-------|-------|-------|
| 1 回目 | ○ | ○ | ○ | ○ |
| 2 回目 | ○ | ○ | ○ | ○ |
| 3 回目 | × | ○ | × | ○ |
| 4 回目 | ○ | ○ | × | ○ |

表 12: 5.3 節を用いた正誤表 1

| 回数 | ユーザ A | ユーザ B | ユーザ C | ユーザ D |
|------|-------|-------|-------|-------|
| 1 回目 | ○ | ○ | ○ | ○ |
| 2 回目 | ○ | ○ | ○ | ○ |
| 3 回目 | ○ | ○ | ○ | ○ |
| 4 回目 | ○ | × | ○ | ○ |

るという点で異なる。また、ユーザの行動範囲をユーザの行動の特徴や興味から推定する研究 [11] も行われている。個別のユーザの興味対象からユーザの行動を予測する点では近いが、本研究では似た趣向を持つ他ユーザの情報も用いる点が異なる。他には、Twitter などのマイクロブログ上に蓄積された人々の行動履歴をもとにユーザの行動とコンテキストを解析することで将来のユーザの行動を予測する研究 [12] なども行われている。本研究とは過去の行動の履歴をもとに将来のユーザ行動を予測する点では通ずるが、Web ブラウザを用いず、位置情報付きのツイートと発言内容から現在地、移動場所を推測する点において、本研究とは大きく主意が異なる。

7. まとめ

本研究では、交通案内サイトの操作・閲覧履歴情報に基づき、最尤法による存在確率と欠損データ補間に基づく行動確率を算出することで、ユーザの将来の移動を推定する手法を提案し、プロトタイプを構築した。ユーザが Web ページに実際に入力した情報だけでなく、過去の入力や他ユーザの履歴により入力された情報を用いることで、欠損データの補間ならびに行動推定の実現の可能性を示すことができた。本手法を用いて、災害等のイベントが発生した際におけるユーザの現在地、または将来移動するであろう場所に、ユーザの状況に適した情報を提示できることを示した。

また、今後の課題として、今回の実験のような目的地となる場所がいくつかの候補の中から選ばれるような場合だけでなく、被験者が自由に選んだ目的地を推定するような実験を行い、手法の有効性をより実際に使用される状況にあわせて検証することや、目的地となる場所にある建物や観光スポットなどの嗜好性を考慮することで、場所同士の類似性に基づいたユーザの移動の推定を行うアイテムベースの協調フィルタリングの実装、災害等の情報提供の実装などが挙げられる。

表 13: 5.3 節を用いた正誤表 2

| 回数 | ユーザ E | ユーザ F | ユーザ G | ユーザ H |
|------|-------|-------|-------|-------|
| 1 回目 | ○ | ○ | ○ | ○ |
| 2 回目 | ○ | ○ | ○ | ○ |
| 3 回目 | ○ | ○ | ○ | ○ |
| 4 回目 | ○ | ○ | ○ | ○ |

謝辞

本研究の一部は、総務省戦略的情報通信研究開発制度 (SCOPE) および JSPS 科研費基盤研究 B(26280042) 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] Leah Hoffmann. Looking back at big data. Communications of the ACM, Vol. 56, Issue 4, pp. 21-23, April 2013.
- [2] 小出 明弘, 斉藤 和巳, 鳥海 不二夫, 風間 一洋, 松尾 真人. デンドログラムを用いた震災時における Mention・Hashtag ツイートの利用傾向分析. WebDB フォーラム 2012
- [3] 細川 侑士, 西村 直也, 白石 優旗, 河合 由起子, 秋山 豊和. ユーザ行動抽出に基づく訪問者に対する災害等イベント情報発信システムの提案, 第 5 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 2013), B5-1, 2013.
- [4] 吉村 直子, 井上 智雄. 東日本大震災前後のマイクロブログサービス Twitter における故郷アカウントの利用分析. 情報処理学会報告, 2012-GN-83(5), 1-8, 2012.
- [5] 山本 修平, 佐藤 哲司. Twitter からの実生活情報の抽出法の提案. 第 4 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 2012), F3-4, 2012.
- [6] 山本 雅人, 小笠原 寛弥, 鈴木 育男, 吉川 正志, 東日本大震災時の Twitter における情報伝播ネットワーク, 情報処理, Vol. 53, No. 11, pp. 1184-1191, 2012.
- [7] 近徳 高, 高井 秀人, 上里 福美. Web サイト訪問者のページ閲覧行動予測システム. 東京工芸大学工学部紀要, Vol. 27 No.1, 2004.
- [8] 大塚 大輔, 久保山 哲二, 安田 浩. Web アクセスログからの閲覧者の行動分析. 電子情報通信学会総合大会講演論文集, D-9-22, 2011.
- [9] 篠田 裕之, 竹内 亨, 寺西 裕一, 春本 要, 下條 真司. 行動履歴に基づく協調フィルタリングによる行動ナビゲーション手法. 情報処理学会研究報告, グループウェアとネットワークサービス, Vol. 2007. No. 91, pp. 87-92, 2007.
- [10] 山田 直治, 磯田佳徳, 南正輝. GPS 搭載携帯電話を用いた移動経路履歴に基づく訪問地・経路地予測システム, 電子情報通信学会技術研究報告 Vol. 110, No. 130, pp. 47-54, 2010.
- [11] 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.
- [12] 斉藤 裕樹, 高山 翼, 山上 慶, 戸辺 義人, 鉄谷 信二. マイクロブログのジオタグと発言コンテキスト解析による行動予測手法. 情報処理学会論文誌, Vol. 55, No. 2, pp. 773-781, 2014.