

Twitter からの眩かれなかった行動の推測手法の提案 ——震災時の帰宅行動に関する事例検討——

グエン ミン テイ[†] 川村 隆浩^{††} 大須賀昭彦^{††}

Proposal of Human Activity Estimation not Tweeted at Twitter ——A Case Study of Returning Home in the Earthquake——

Nguyen MINH THE[†], Takahiro KAWAMURA^{††}, and Akihiko OHSUGA^{††}

あらまし 本論文は、実世界での人々の行動を分析し、状況に応じた適切な情報提示などへ役立てることを目的に、Twitter などソーシャルメディアから行動ネットワークを構築する研究の一環である。ソーシャルメディアからの行動抽出にあたっては、さまざまな理由から眩かれなかった行動が数多く存在し、結果として行動ネットワークがスパースになってしまうという問題が存在する。そこで、行動の性質とユーザのゴールを考慮した行動ベース協調フィルタリング手法を提案し、欠損行動の推測を試みる。また、協調フィルタリングによる頻度の原理の副作用としての低頻度だが価値のある情報が埋もれてしまう問題に対して、人が成功している人や行動にどのように影響を受けるか、を単純にモデル化し、一定の重み付けを行う方法を提案する。そして、東日本大震災発生時の tweet 337,958 件を対象に評価実験を行った結果、提案手法を用いることで行動ネットワーク内の欠損行動ノードを一定程度、補完できることを確認した。

キーワード Web マイニング, 協調フィルタリング, 実世界指向エージェント

1. ま え が き

近年、実世界での人々の行動に基いて、スマートフォン等を通じて適切な情報を提供するサービスが盛んである。我々は従来より実世界指向エージェント研究の一環として、ブログから人々の行動を分析し、マーケティング等へ活かす研究を進めてきた [1]。しかし、東日本大震災 (3.11) 後は Twitter などのソーシャルメディアを対象に、よりリアルタイムで細粒度な行動把握を指向し、避難情報の提示などへ役立てる研究を進めている [2]~[5]。従来研究では、人々の行動を Linked Data として行動ネットワーク化することを目的に、CRF(Conditional Random Fields) [6] を用いてソーシャルメディアから RDF(Resource Description Framework) [7] の基本要素であるトリプルの抽

出を試み、一定の精度で抽出可能であることを示した [5]。これにより行政側による市民行動把握や、それに基づく適切な避難方法の指示、救援物資の支援、更には復興時の避難経路再設計等への応用を目指している (人々が直接利用するサービスではなく、行政機関等が全体を把握するためのサービスの位置付け)。具体的には、多くの人が電車移動を試みた後、何らかの理由で諦めて歩き出したことが分かれば、電車の運行状況に関する情報を人々に広く周知したり、場合によっては推測された行動が危険なものであれば (十分な高さのない場所への避難など)、津波等の危険を伝えてより高台への避難を促すといった利用を考えている。災害発生後の街づくりの際にも、住民行動のフローとボリュームを把握することは街の動線設計に役立つだろう。

ここで行動ネットワークとは、行動間の関係を表すネットワーク (有向グラフ) である。ノードは行動の構成要素 (動作, 対象, 時間, 場所), リンクはそれらの関係 (遷移関係 next, 因果関係 becauseOf など) を表す。行動ネットワークの例を図 1 に示す。図中、

[†] 楽天株式会社, 東京都

Rakuten, Inc., Shinagawa-ku, Tokyo, 140-0002 Japan

^{††} 電気通信大学大学院情報システム学研究所, 調布市
Graduate School of Information Systems, The University of
Electro-Communications, Chofu-shi, 182-8585 Japan

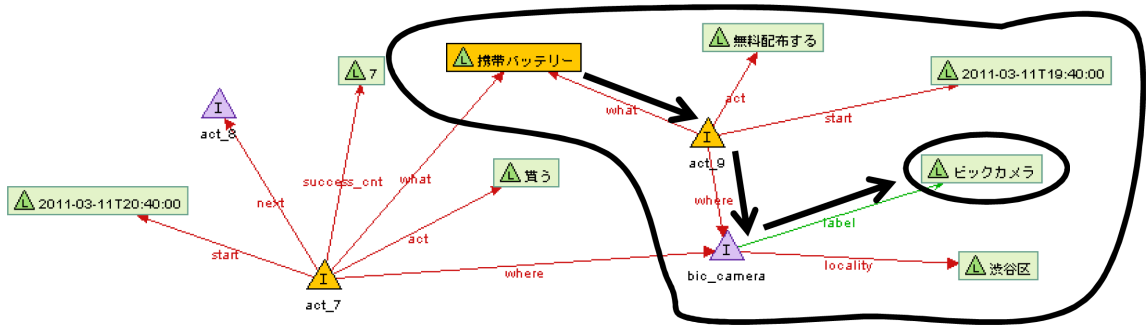


図1 行動ネットワークの一部 (I, L はそれぞれインスタンス, ラベルを表す)
 Fig.1 Part of activity network. (I, L means Instance, Label, respectively)

携帯バッテリーのラベルからリンク (what) を辿り、ノード (act_9) を介して、場所 (where) を参照すると、ビックカメラが携帯バッテリーを配布していたことが読み取れる。同様に、交通機関の運行状況やトイレ・避難所の情報も得られることを確認した。我々はこうした情報を用いて、ユーザ状況に応じて適切な情報を提供するエージェントサービスの実現を目指している。

しかし、上記 [5] でも言及しているとおり、ソーシャルメディアはユーザの自主性に依存しているため、眩れなかった、または何らかの事情により眩れなかった行動も数多く存在する。結果として、ノードが欠損し、行動ネットワークがスパースになってしまうという問題点が存在する。したがって、住民行動のフローや各フローのボリュームをより正確に把握するためには、実 tweet で得られた情報に加えて、欠損行動を推測する必要がある。そこで、本論文では協調フィルタリングの原理を用いて欠損行動の補完を試みる。また、頻度の原理の副作用ともいえる、低頻度だが価値のある情報が埋もれてしまう問題に対して、人が成功している人やその行動にどのように影響を受けるかをモデル化することで補正する手法を提案する。

本論文の構成は次のとおりである。2. では、欠損行動を補完するため、行動ベース協調フィルタリング手法を提案する。3. では、特に低頻度の成功行動を補正するため、上記手法の改良を提案する。4. では、実際の tweet を対象に本提案手法の有効性を評価する。最後に、5. で関連研究を示し、6. でまとめと今後の課題について述べる。なお、tweet 文からの行動ネットワーク構築に関する詳細は [1], [3] を参照して頂きたい。

2. 行動ベース協調フィルタリング手法による欠損行動の補完

任意の時刻におけるユーザの欠損行動を推測する問題は、時刻 t にユーザ $u_i (i = 1, \dots, L)$ が場所 $l_n (n = 1, \dots, N)$ で行動 $act_m (m = 1, \dots, M)$ を行う可能性を計算する問題といえる。これを $P(u_i, act_m, l_n, t)$ と表す。ただし、以下では単純化のため、推測対象となるユーザ u_i が時刻 t に行ったかもしれない行動を act と表記する。同様に、ユーザ u_i が時刻 t に存在していたと思われる場所を l と表記する。この問題に対し、我々は協調フィルタリング手法におけるアイテムベースフィルタリングとユーザベースフィルタリングを組み合わせることで算出を試みた (図 2)。本問題に協調フィルタリングの応用を考えついた発端は、人々の行動はたぶん右に倣え的であり、特に震災時の避難といった状況においてはその傾向が強く、アイテムやユーザの類似度に基づいて推薦するという協調フィルタリングの考え方に沿って定式化しやすいと考えたためである。ただし、協調フィルタリング以外でも、例えばリンクマイニングにおけるリンク予測問題として、以下の人間行動に関する仮定をノード間の周辺リンク構造に基づいて定義される構造情報に基づく特徴ベクトルとして表現する指標を考案することで、教師付き学習問題として定式化できるだろう。また、HMM であれば、状態と状態遷移確率で表現する手法を考案することで対応できるだろう。しかし、著者らの調べた限りでは、これらいずれの方法においても本論文で扱っている欠損行動の補完に適用した事例は報告されていない。

従来の協調フィルタリングの応用は、ユーザの嗜好

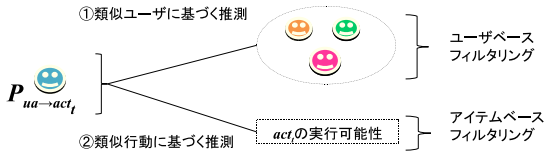


図2 欠損行動を推測するアプローチ
Fig.2 Approach to estimate missing activities.

に合うアイテム (EC サイト内の商品など) を推薦することであるのに対し, 本論文の目的は欠損行動を推測することである^(注1) (表1に, 行動とアイテムの性質の違いを示す). そこで, 我々は協調フィルタリング手法における代表的なメモリベース法の一つ, GroupLens [8] を参考に, 行動の性質を考慮した協調フィルタリング手法を提案する. 具体的には, 評価値行列を作成して行または列ベクトル間の類似度に基づいて推測するという点は共通だが, データの性質上, 各変数と場所, 遷移関係, ゴールなどに強い依存関係が存在するため, 以降に示す行動ベースの相関関数を新たに定義する.

2.1 ユーザの類似度に基づく推測

まず我々は, ユーザ間の類似度から欠損行動を補完することを試みる. ここでは以下の考え方を前提としている.

- 類似ユーザは同様の前後行動を行う可能性が高い.
- 同様のゴール (避難したい, 帰宅したい等) をもつユーザは同様の行動を行う可能性が高い.
- 震災時に, 一定範囲内にいるユーザは同様の行動を行う可能性が高い.

このことから, ユーザ間の類似度を式 (1) に示す.

$$S(u_i, u_j) = \gamma \text{SameActions}(u_i, u_j) + \delta \text{SameTarget}(u_i, u_j) \quad (1)$$

ここで SameActions は, ユーザ u_j がユーザ u_i と同じ場所で同じ前後行動を行ったか否かを計る関数である. 具体的には, ユーザ u_i が推測対象行動 act の前に行った行動 act_{before} と後に行った行動 act_{after} , 前に居た場所 l_{before} と後に居た場所 l_{after} が, ユーザ u_j のそれと同一である場合は 1, 異なる場合は 0 とする. SameTarget は, ユーザ u_j がユーザ u_i と同じゴールをもつかどうかを判断する関数である. 同じ場合は

(注1): 協調フィルタリングというアイテムへの評価値の欠損とは異なる

表1 アイテムと行動の比較
Table 1 Comparison of item and activity.

比較観点		EC サイト のアイテム	震災時の行動 四つの変数 (動作, 対象, 時刻, 場所)
複雑さ		一つの変数	
依存関係	場所	NO	YES
	遷移関係	弱い	強い
	ゴール概念	NO	YES (避難, 等)
実行時間の考慮		不要	必要

1 になり, 異なる場合は 0 とする. ゴールの判別については 4.1 を参照して欲しい. なお, 時刻 t の開始時間や単位時間はアプリケーションごとに設定するものとし, それに基づいて前後の判定を行う. γ, δ は係数であり, $0 \leq \gamma, \delta, \gamma + \delta \leq 1$ を満たすものとする.

2.2 行動の連続性に基づく推測

次に, 実世界の行動は前後の行動, 時間, 場所, 行動間の流れに依存するため, これらの要素を考慮する. 行動 act が起こる可能性 $C(act, t, l)$ を式 (2) で表す.

$$C(act, t, l) = \rho_a (F_{before}(act) + F_{after}(act)) + \rho_t F(act, t) + (1 - \rho_a - \rho_t) F(act, l) \quad (2)$$

ここで, $F_{before}(act)$ は, その時点での行動ネットワーク全体を通じて act に先立つ全行動の内 (遷移関係リンク $next$ の出し元), act_{before} が現れる割合である. 同様に, $F_{after}(act)$ は act に続く全行動の内 (遷移関係リンク $next$ の指し先), act_{after} が現れる割合である. なお, 前節と同様に act_{before} はユーザ u_i が act の前に行った行動, act_{after} は後に行った行動を表す. また, $F(act, t)$ は時刻 t における行動 act の頻度, $F(act, l)$ は場所 l における行動 act の頻度を表す. ρ_a, ρ_t は係数であり, $0 \leq \rho_a, \rho_t, \rho_a + \rho_t \leq 1$ を満たすものとする.

2.3 実行時間に基づく推測

更に, 対象ユーザ u_i は眩かなかった時間帯に複数の行動を行った可能性もある. 例えば, 図3では, 対象ユーザは眩かなかった時間帯に act_{before} (眩いた行動) を除き, act_1 と act_2 を行った可能性がある. このため, 対象ユーザ u_i は眩かなかった時間内の任意の時刻 t に, act_{before} の実行中であったか, または act_1 または act_2 のいずれを行っていたのかを推測する必要がある. そのため, ここでは類似ユーザ u_j の行動とその実行時間を考慮する. 類似ユーザ u_j が行った

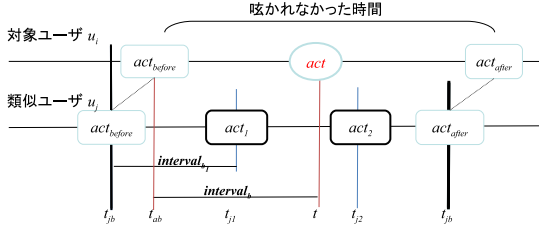


図 3 眩れなかった時間に複数の行動が行われた場合
 Fig. 3 Existence of more than two activities during not-tweeted period.

$act_{before}, act_1, act_2$ の開始時刻をそれぞれ t_{jb}, t_{j1}, t_{j2} とすると, act_{before} の実行時間 $interval_{b_1}$ は $[t_{jb}, t_{j1}]$ である. 更に, $[t_{ab}, t]$ までの時間を $interval_b$ とすると, act_1 が行われた可能性を以下のように評価できる.

$$T(act_1, u_i, u_j) = \begin{cases} \frac{interval_b}{interval_{b_1}} & \text{if } diff_time < 0 \\ \frac{1}{diff_time+1} & \text{else} \end{cases} \quad (3)$$

$$diff_time = interval_b - interval_{b_1} \quad (4)$$

つまり, $interval_{b_1}$ と $interval_b$ が同じ値の場合, act_1 を行った可能性を最も高いものとする. そして, $interval_{b_1}$ と $interval_b$ の差が大きくなるほど, act_1 を行った可能性を減少させている.

2.4 欠損行動の推測式

上述した式 (1), (2), (3) を合わせることで, 推測評価値 $P(u_i, act, l, t)$ を式 (5) のように求めることができる.

$$P(u_i, act, l, t) = \alpha \left(\frac{\sum_{j=1, \dots, L} \omega(u_j, act) * S(u_i, u_j) * T(act, u_i, u_j)}{L} \right) + (1 - \alpha)C(act, t, l) \quad (5)$$

なお, L はユーザの数である. $\omega(u_j, act)$ は, 行動 act に対するユーザ u_j の重みである. ユーザ u_j が act を行った場合は 1 になり, 行わなかった場合は 0 とする. α は係数であり, $0 \leq \alpha \leq 1$ を満たすものとする. 解決したい課題や行動データによって, これらの値を決めるものとする.

3. 成功行動の影響を考慮した低頻度情報の補正

前述した協調フィルタリング手法による, いわば多数決の副作用として低頻度だが価値のある行動 (結果として成功した行動. 以下, 成功行動と呼ぶ) が埋もれてしまい, 行動ネットワーク上に再現されないという問題がある. 単純な例として, 同様の失敗行動が三つ, 別の成功行動が二つあり, 内一つの成功行動が欠損してしまった場合, 前述の方法では失敗行動として再現される可能性が高い. これは, 冒頭で述べた行動ネットワークの構築目的に照らして望ましくない. 災害時には, 低頻度だが成功した行動については適切に把握し, 必要に応じて行政側から市民へのフィードバックを行うことに利用したい. また主観的には, 人は頻度は低くても成功した例は目敏く見つけ出して真似る, という特性があるようにも思われる. そのため, 必ずしも右に倣えの協調フィルタリングによる補正に従わないことも多い. しかし, 基本的に協調フィルタリングは高頻度のパターンに基づくため, マジョリティ情報は容易に扱えるが, マイノリティ情報に対応するのは困難である (被覆率の観点で不可能という意味ではない. セレンディピティに近いと考えている). そこで, 人が成功している人や行動にどのように影響を受けるか, を単純にモデル化し, 他人の成功が与えたかもしれない影響を加算する手法を提案するなお, 低頻度かつ成功率の高い行動を, 協調フィルタリングの計算上, 高頻度情報に左右されずに補完する処理を補正と呼んでいる.

反対に, 成功例の方が数が多く, 失敗例が再現されない可能性もあるが, これは上記目的に照らして無視してよいとする. ただし, ここではオープンワールドを仮定しているため, 成功例だけを抜き出せばよいとは考えていない. 行動ネットワークに現れていない行動は unknown であり, 失敗と分かっている行動 (negative) はネットワーク上に陽に現すことが必要と考えている.

(1) 言語処理で成功行動を抽出

まず言語処理 (品詞解析, 構文解析) 手法を用いて, “して良かった”, “が正解だった” など成功を表す言葉と係り受け関係をもつ行動を成功行動とみなし, tweet 文中に現れる成功行動を抽出する. 図 4 に例を示す.

(2) 対象ユーザが成功行動を行った可能性を推測

次に, 以下の単純な考え方を前提として, 対象ユー

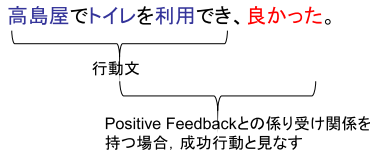


図 4 言語処理で成功行動を抽出
Fig. 4 Extraction of successful activity using NLP

表 2 u_i の行動例
Table 2 Example activities.

行動列	成功行動	行動ごとの 成功回数	時間
act_{before}	F	0	t_{before}
act_1	F	0	t_1
act_2	T	5	t_2
act_3	T	1	t_3
act_{after}	F	0	t_{after}

が成功行動を行った可能性を推測する。

- よく成功している人 (成功行動の回数が多い人) は、成功する可能性が高い (式 (6))

- 人はより多くの人々が成功している行動 (成功回数が多い行動) を選ぶ可能性が高い (式 (7))

これにより、対象ユーザ u_i が成功行動を行った可能性を式 (8) で表す。

$$f(u_i) = \frac{\text{成功した行動の数}}{u_i \text{の行動の数}} \quad (6)$$

$$Success(act) = \frac{act \text{の成功回数}}{\text{全ての行動の成功回数の和}} \quad (7)$$

$$DidSuccess(u_i, act) = f(u_i) * Success(act) \quad (8)$$

例として、 $t_{before} \sim t_{after}$ 間の u_i の行動が表 2 のとおりであった場合、式 (6) と式 (7) の値はそれぞれ以下ようになる。

$$f(u_i) = \frac{2}{5} \quad (9)$$

$$Success(act_2) = \frac{5}{5+1} \quad (10)$$

(3) 協調フィルタリングとの統合

上記に沿って成功行動を考慮し、最終的に対象ユーザ u_i による行動 act の推測評価値を以下のように計算する。なお、 β は両者のスケールを調整するための係数である。

$$P^*(u_i, act, l, t) = P(u_i, act, l, t) + \beta DidSuccess(u_i, act) \quad (11)$$

4. 評価実験

本章では、上記行動ベース協調フィルタリングと言語処理を用いた欠損行動の補完実験の結果を示す。

対象データは、3.11 発生時の避難に関する tweet 337,958 件、2,907 名分から構築された行動ネットワークである。tweet の内容の多くは、不特定多数への情報提供や友人、知人に向けて自分の行動、状況を知らせる内容であった。今回、行動ネットワーク構築に用いたシステムの tweet から行動ネットワークへの変換精度はノードの種類によって 70-90% であった [1]。変換精度が低い場合は欠損行動を推測するよりも、書かれている内容の解析に当たった方が有意義と思われるが、一定程度の精度は出ていることと、行動ネットワーク構築の研究は現在も継続しており精度は日々向上していることから、本論文では問題の切り分けのため、正しく変換できたものを対象とする。また、行動ネットワークは 4 種類 (動作, 対象, 時間, 場所) のノードから構成されるが、ここでは動作ノードを推測対象とした。動作ノードは、別途構築された避難オントロジー [2] に基づいて、以下、15 種類のインスタンスのいずれかをもつものとする {歩く, 電車で移動する, 車・バス・タクシーで移動する, 休憩・待機する, 帰宅する (移動手段不明), 充電する, トイレに行く, 声掛け・会話する, 物品を提供する, 避難所情報の提供, その他情報の提供, 犯罪情報の提供, 家族・知人に連絡する, 医療機関を呼ぶ, その他}。クラス、インスタンスの構成については本論文では省略したい。なお、提案手法の適用条件は、行動を伴う 2 以上のユーザ、行動の時刻と位置情報、推定対象行動に next リンクで繋がる前後行動及びゴール行動が行動ネットワーク上に表れており、元 tweet から行動の成否が判別可能であることとする。

4.1 欠損行動の補完

実験では、行動ネットワークからランダムに動作ノード act を削除、欠損行動とし、これを式 (5) に基づいて推測した。具体的には誤差範囲を 0.5% と設定し、平均値の変動が 0.5% を切ったところで終了した。実際には、いずれの場合も 100 回以下で終了している。なお、補完対象ユーザ u_i 、時刻 t 、 act_{before} 、 act_{after} は既知とする。つまり、 $act_{before} \rightarrow ?act \rightarrow act_{after}$ の中

から、 $?act$ を当てる問題である (\rightarrow は、状態遷移を表す next リンクである). 行動の時刻は、頻度計算の関係上、あまり細かく分けても意味がないが、時間が経てば状況が変わることから、tweet のタイムスタンプから取得して 1 時間を一単位とした. なお、実験によれば 1000 行動を 1.5 秒程度でネットワーク化できることを確かめており、地域を限定すればリアルタイムでも、ある人が取った過去 3 時間の行動から別の人の欠損行動を補完することができるだろう (ただし、ゴール判別には仮定を入れる必要がある). また、場所は新宿に限定した. なお、対象データ中、新宿に居たと思われる人 (ただし、新宿に関する情報を発信している人、実際に新宿にいたかどうかは不明を含む) は 1,165 名、165,363 行動 (動作ノード “その他” を除く) であった. 動作ノード別の行動数、及び成功行動数 (括弧内は割合) を表 3 に示す.

ユーザのゴール判別に関しては、ここではゴールを帰宅に統一し、あきらかに関係のないものを取り除いた. ユーザの行動全てにゴールがあるわけではないため、この点はアプリケーションに依存する条件となるが、他には電車事故時に出勤する、バーゲンで買い物をする、アイドルのコンサート行く、など群衆がある目的に向かって一斉に行動する場面への応用を想定している. また、ゴール判別は今回のようにオフラインであれば解析できるが、オンライン (イベント発生時にリアルタイムの解析) では難しいという問題もある. この点に関しては、行動連鎖の途中でゴールが判別できればネットワークに加えるなど、使い方を工夫する必要があるだろう. 表 4 に実験結果を示す. 再現率は正しく再現された行動数の割合を表す. 分母は、各動作ノードの行動数である.

実験の結果、行動ネットワーク内で比較的、出現頻度の高い行動「電車で移動する」の再現率は、80%を超える結果を得た. 一方で、相対的に頻度の低い「休憩・待機する」の再現率は 50%を少し超える程度であった. また、ランダムにサンプルを拾った場合 (サンプルは行動ネットワーク内の行動頻度と同じ分布で得られるものとする) に相当する全体の再現率 (各行動の再現率を頻度割合に応じて合算したもの) は 80%弱という結果を得た.

なお、係数は実験を通して調整し、 $\gamma = 0.33, \delta = 0.33, \rho_a = 0.8, \rho_t = 0.1, \alpha = 0.2$ と設定した. 行動の頻度に掛かる ρ_t を大きくすると、高頻度の行動の再現率は 100%近くに上がるが (適合率が下がり)、全体

表 3 動作ノード別の行動数
Table 3 Number of activities sorted by type.

動作ノードの種類	行動数	成功行動数 (%)
その他情報の提供	38696	110 (0.28)
避難所情報の提供	32081	86 (0.27)
家族・知人に連絡する	30878	243 (0.79)
車、バス、タクシーで移動する	10836	62 (0.57)
医療機関を呼ぶ	9474	7 (0.07)
物品を提供する	8714	10 (0.11)
電車で移動する	8524	36 (0.42)
帰宅する	7395	126 (1.70)
犯罪情報の提供	4697	16 (0.34)
休憩・待機する	4354	41 (0.94)
充電する	4025	13 (0.32)
トイレに行く	2260	2 (0.09)
声掛け、会話する	1883	5 (0.27)
歩く	1547	33 (2.13)
合計	165363	790 (0.48)

表 4 欠損行動の補完
Table 4 Complement of missing activity.

対象行動	電車で移動する	休憩・待機する	全行動
再現率	81.8%	54.5%	77.5%
推測評価値 P (平均)	0.481	0.414	-

としては低下する. そこで、推測対象行動単体の頻度よりも前後行動のつながりを重視し、行動の連続性に掛かる ρ_a を大きくし、最終的に掛かる $(1 - \alpha)$ も大きくすることで、低頻度の行動とのバランスから結果的に全体が向上することを確かめた. ただし、これは行動の種類が多寡 (ここでは 15 種類) が影響し、行動の種類が少ない場合は推測対象行動固有の前後行動のつながりが得られないことが予想される.

この結果から、提案手法は 8 割近くのノードを再現できることを確認した. しかし、精度は最終的には乗数となっていくため、2 ノードの組合せでは $77.5\%^2 \cong 60\%$ といった具合に低下するだろう. そのため、欠損がピンポイントであった場合は一定の補完が可能だが、一定時間にわたって tweet が無いなど、欠損行動が連続していると思われる場合には有効性は半減すると予想される (注2).

4.2 低頻度情報の補正

前節と同様に、低頻度であるが価値のある行動 (成功行動) の補正効果を確認するため、式 (11) に基づいて動作ノード act を推測した. 先の実験結果に挙げた二つの行動の内、「電車移動」は高頻度であるが成功例は少なく、相対的に低頻度である「休憩」の方が成

(注2): ただし、災害時の避難支援という観点では、半分でも推測できるならば役に立つ場面もあるだろうと考えている.

功とみなされる割合が大きかった。これは、「電車が止まっており、結局、オフィスに帰ってきた」や「避難所で食事と暖を取れて良かった」など、一般に言われていることに対応している。内閣府（防災担当）の資料 [9] でも、むやみに移動せず、休憩・待機を勧めており、行動ネットワークにおいてはこれをより正しく推測したいと考えた。表 5 に実験結果を示す。

実験の結果、成功行動である「休憩」の再現率が 10% 程度、向上できたことを確認した。しかし、高頻度行動の再現率が低下し、全体の内、多くの割合を占める高頻度行動が下がったことで、全体では約 7% 下がった。個々のケースを見てみると、先の実験において推測評価値 P が非常に僅差で負けていたものが（表 4 から両者の差は平均的にも大きくないことが分かる）、式 (11) の第 2 項によって逆転している。

「電車」、「休憩」以外の行動では、成功行動を補正しても高頻度のものが推測しやすく（再現率が高く）、低頻度のものは推測しにくい（再現率が低い）という全体傾向は変わらなかった。行動数最少の「歩く」で再現率約 50%、最多の「その他情報の提供」では 90% 超であった。無論、この傾向は前節で考察した行動の種類の多寡や、遷移関係などの行動ネットワークの状況に依存するが、今回、高頻度トップ 3 であった情報提供や家族への連絡はあらゆる行動との遷移関係が見られ、純粋に出現頻度が大きく左右した形となっている。なお、本論文において「電車」と「休憩」を例に考察したのは、この二つが表 3 の中で比較的中間に位置し、かつ、行動数は「電車」が「休憩」の約 2 倍、成功率は「休憩」が「電車」の約 2 倍と好対照を示しており、頻度と補正の関係を把握しやすいと考えたためである。また、併せて内閣府資料における定性的な分析との対応を確認するためである。

なお、係数は先の実験と同様とし、スケーリング係数 β は 100 に設定した。この値を大きくすると、低頻度の行動の再現率をより上げることができるが、高頻度の行動とのバランスから実験を通して調整した。

また、式 (7) における成功回数は、それぞれ同一の

場所（新宿）での対象行動の成功回数と、全ての成功した行動の回数とした。成功かどうかの判定には約 20 のルールを作成して判別した。代表的なものとしては、構文解析の結果、行動を表す句が positive word（～して良かった、～は正解だったなど）に掛かっているものや、逆に positive word から行動に掛かっているもの（幸いしたのは～したことだ、うまく～できた）、または行動を表す文の後に感謝の言葉が続くものなどが挙げられる。更に、上記のバリエーションとして negative word の否定表現や、2 重否定なども扱っている。なお、全ての成功行動回数は 790 件（表 3 参照）であった。

5. 関連研究

本章では、まず協調フィルタリングに関する関連研究について述べる。協調フィルタリングには、大きく分けてアイテムをベースとする手法（アイテムベース）とユーザをベースとする手法（ユーザベース）が存在する。アイテムベースとは、他のアイテムとの類似度とそのアイテムの評価値を用いて推薦対象ユーザの未評価値を推測する手法である。またユーザベースとは、他のユーザとの類似度とそのユーザの評価値を用いて推薦対象のユーザの未評価値を推測する手法である。本提案手法は、行動を推測するためにこれらを組み合わせ、更に時間と場所の関連性を考慮した形になっている。Ma ら [10] は、やはりアイテムベースとユーザベースを統合した手法を提案しているが、時間や場所は考慮していない。また、Koren [11] はアイテムベースとユーザベースを統合した上で時間も考慮した手法を提案しているが、目的の違いから場所は考慮していない。更に、川前ら [12] は履歴の選別にユーザごとの嗜好、及びユーザ間の関係の時系列変化の両方を反映させた協調フィルタリング手法を提案している。これはユーザ間でのアイテム購入の先行関係を用いた協調フィルタリング手法であり、「各ユーザは同様の嗜好をもつ先行ユーザと同様の行動をとり、先行性は時間変化する」という仮説に基づいている。そして、先行度の計算にユーザの嗜好の先行性を反映するため、ユーザ間の購入時刻の差を導入し、更にユーザごとの履歴の時間減衰を導入している。これらの考え方は参考にすべき点も多く、本提案手法の今後の改良に活かしていきたいと考えているが、この手法も上記と同様に場所の概念は考慮していない。また、いずれも低頻度情報の補正といった考え方は示していない。

表 5 欠損行動の補完（低頻度情報補正付き）
Table 5 Complement of missing activity with minor info. correction.

対象行動	電車で移動する	休憩・待機する	全行動
再現率	72.7%	63.6%	70.7%
推測評価値			
P^* (平均)	0.553	0.733	-

次に、震災時避難といった適用ドメインから関連研究について述べる。グーグル株式会社らが主催し、2012年9,10月に開催された「東日本大震災ビッグデータワークショップ - Project 311 -」[13]では、Twitterのデータを扱ったものが数多く見られた。鳥海ら[14]のプロジェクト CHIDRI は、災害救助支援を実現するための基盤技術を構築することを目的とした人工知能学会の近未来チャレンジテーマであるが、平均連続沈黙時間（最後に投稿してから経過した時間のユーザ平均）と被災状況との相関性を求めるなど、本論文の目的である眩れなかったことの推測とも関連があり、大変興味深い。また、石野ら[15]は、Twitterから行動経路に特化して抽出を試みており、その信憑性の検証を試みているが、いずれも行動ネットワークの構築や、その欠損情報の補完を扱ったものではない。

6. む す び

本研究では、災害時にユーザ状況に応じて適切な情報を提供するエージェントシステムの実現を目指し、Twitter上の情報からユーザの行動を時系列的に表した行動ネットワークを構築している。特に、本論文では行動ネットワーク構築にあたって眩れなかった行動（欠損行動）を推測するために、行動の性質とユーザのゴール（避難、帰宅など）を考慮した行動ベース協調フィルタリング手法を提案した。また、成功が人に与える影響をモデル化することで、低頻度であるが価値のある行動（成功行動）に一定の重み付けを行うことを提案した。そして、東日本大震災発生時のtweet 337,958件を対象に評価実験を行った結果、提案手法を用いることで行動ネットワーク内の欠損行動ノードを一定程度、補完できることを確認した。また、成功行動のモデル化についても低頻度行動の推定精度を向上できたことから一定の妥当性があると考えている。しかし同時に、同様な前後行動をもつ高頻度行動の精度低下を招いており、行動ベース協調フィルタリングとの組み合わせ方法に課題を残した。恐らく各行動ごとに成功とそれ以外の確率分布を求めて、組み合わせに活かす処理が必要になるものと思われる。この点については今後の課題としたい。

また、場所情報と行動とを関連付けることで欠損行動の推測に役立てることができただろう（駅のないところで電車には乗れない、など）。現状、tweetに付随するGEOタグ（GPS情報）は数少なく、また必ずしも正確ではないこと、そのためtweetの内容から場所

に関する単語を抽出して位置情報を追加していることなどから、住所やPOIといったレベルの位置情報は取れておらず、新宿や渋谷といった範囲でのみ扱っている。しかし、今後、より詳細な位置情報が取得できた際には推測精度の改善に活用していきたい。

なお、今回は過去の欠損行動の推測を行っているが、今後の課題としては将来行う可能性がある行動の推測への応用も検討して行きたい。また、さまざまな自然災害（台風、大雪）、人的災害（ストライキ、テロ）時に人々の行動把握・情報共有に役立てることを目指して、行動ネットワークに基づく実世界指向エージェントサービスの立ち上げを進めたい。更に、いつ起こるとも知れない災害時に迅速に活用するためには、平常時から本サービスを一般に使ってもらうことが重要だと考えている。そのため、街中での観光やマーケティングといった用途での活用も併せて検討して行きたい。

謝辞 本研究の一部は、科研費(B)「ミニブログを利用した災害時避難行動推薦システムのためのクラウドミドルウェアの研究」(24300005)の助成を受けたものである。

文 献

- [1] ゲン ミン ティ, 川村隆浩, 中川博之, 田原康之, 大須賀昭彦, “条件付確率場と自己教師あり学習を用いた行動属性の自動抽出と評価,” 人工知能学会論文誌, vol.26, no.1, pp.166-178, 2011.
- [2] ミヤマエ イワナガ イサベル シズ, ゲン ミン ティ, 川村隆浩, 中川博之, 田原康之, 大須賀昭彦, “震災時の効率的な避難のための行動推薦エージェント Ready for 87%の提案 (1) - 避難計画オントロジーの設計,” 合同エージェントワークショップ&シンポジウム (JAWS 2011) 論文集, 2011.
- [3] ゲン ミン ティ, 越川兼地, 川村隆浩, 田原康之, 大須賀昭彦, “震災時の効率的な避難のための行動推薦エージェント Ready for 87%の提案 (2) - 時系列避難行動ネットワークの構築,” 合同エージェントワークショップ&シンポジウム (JAWS 2011) 論文集, 2011.
- [4] 伊藤雅博, ゲン ミン ティ, 川村隆浩, 中川博之, 田原康之, 大須賀昭彦, “震災時の効率的な避難のための行動推薦エージェント Ready for 87%の提案 (3) - 行動ネットワークを用いた避難行動の推薦,” 合同エージェントワークショップ&シンポジウム (JAWS 2011) 論文集, 2011.
- [5] T.M. Nguyen, T. Kawamura, Y. Tahara and A. Ohsuga, “Building a time series action network for earthquake disaster,” Proc. International Conference on Agents and Artificial Intelligence (ICAART2012), 2012.
- [6] J. Lafferty, A. McCallum, and F. Pereira, “Conditional random fields: Probabilistic models for segmenting and labeling sequence data,” Proc. ICML2001, 2001.

- [7] RDF: <http://www.w3.org/TR/rdf-concepts/>
- [8] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: An open architecture for collaborative filtering of netnews," Proc. Conference on Computer Supported Cooperative Work, 1994.
- [9] 内閣府 (防災担当): 震災時の帰宅行動, そのときあなたは どうする?, http://www.bousai.go.jp/jishin/chubou/taisaku_syuto/pdf/101110/101110.pdf
- [10] H. Ma, I. King, and M.R. Lyu, "Effective missing data prediction for collaborative filtering," Proc. SIGIR, 2007.
- [11] Y. Koren, "Collaborative filtering with temporal dynamics, Proc. KDD, 2009.
- [12] 川前徳章, 坂野 鋭, 山田武士, 上田修功, "ユーザの嗜好の時系列性と先行性に着目した協調フィルタリング," 信学論 (D), vol.J92-D, no.6, pp.767-776, June 2009.
- [13] 東日本大震災ビッグデータワークショップ - Project 311 -: <https://sites.google.com/site/prj311/>
- [14] 異種協調型災害情報支援システム実現に向けた基盤技術の構築 (CHIDRI): <https://sites.google.com/site/crepchidri/home>
- [15] Twitter からの被災時の行動経路の自動抽出およびその信憑性の検証: <https://sites.google.com/site/prj311/project/twitter-karano-bei-zai-shino-xing-dong-jing-luno-zi-dong-chou-chuoyobisono-xin-ping-xingno-jian-zheng>

(平成 25 年 2 月 8 日受付, 6 月 17 日再受付)



大須賀昭彦 (正員)

1981 年上智大学理工学部数学科卒。同年, (株) 東芝入社。同社研究開発センターソフトウェア技術センター等に所属。1985-1989 年 (財) 新世代コンピュータ技術開発機構 (ICOT) 出向。2007 年より電気通信大学大学院情報システム学研究科教授。2012 年より国立情報学研究所客員教授兼任。工学博士 (早稲田大学)。主としてソフトウェアのためのフォーマルメソッド, エージェント技術の研究に従事。1986 年度情報処理学会論文賞受賞。IEEE Computer Society Japan Chapter Chair, 人工知能学会理事, 日本ソフトウェア科学会理事を歴任。情報処理学会, 電子情報通信学会, 人工知能学会, 日本ソフトウェア科学会, IEEE CS 各会員。



グエン ミン テイ

2012 年電気通信大学大学院情報システム学研究科社会知能情報学博士課程了。博士 (情報学)。同年, 楽天株式会社入社。現在に至る。



川村 隆浩

1994 年早稲田大学大学院理工学研究科電気工学専攻修士課程了。同年, (株) 東芝入社。現在, 同社研究開発センター主任研究員。工学博士 (早稲田大学)。2001-2002 年米国カーネギー・メロン大学ロボット工学研究所客員研究員。2003 年より電気通信大学大学院情報システム学研究科客員准教授。2007 年より大阪大学大学院工学研究科非常勤講師。主としてマルチエージェントシステム, セマンティック Web の研究・開発に従事。人工知能学会, 情報処理学会各会員。