法政大学学術機関リポジトリ
HOSEI UNIVERSITY REPOSITORY

PDF issue: 2025-01-15

ライフログ分析による行動特徴抽出及びイベント検出

角田, 宏貴 / SUMIDA, Hiroki

(出版者 / Publisher)
法政大学大学院情報科学研究科
(雑誌名 / Journal or Publication Title)
法政大学大学院紀要. 情報科学研究科編 / 法政大学大学院紀要. 情報科学研究科編
(巻 / Volume)
9
(開始ページ / Start Page)
119
(終了ページ / End Page)
124
(発行年 / Year)
2014-03
(URL)
https://doi.org/10.15002/00010532

ライフログ分析による行動特徴抽出及びイベント検出

The Lifelog Analyses for Activity Pattern Extraction and Event Detection

角田　宏貴

法政大学大学院情報科学研究科情報科学専攻

E-mail: hiroki.sumida.2k@cis.k.hosei.ac.jp

Abstract—Recently, obtaining lifelog becomes much easier and they can be acquired at any time, thanks to the explosive spread of smartphones and the downsized wearable logging devices. From lifelog it is possible to examine user's activities and life styles. In a smartphone there are various kinds of embedded sensors, e.g., GPS, acceleration sensor, luminance sensors and so on, that can be used to capture lifelog. These sensors make smartphones as a valuable life logging tool. In addition, smart phones can be connected to the Internet and this makes it possible to obtain logs when using Web applications, for instance, Web browsing histories and SNS logs like Twitter. There are a lot of researches for analyzing smartphone logs, for examples, the extraction of user's activity using location, date and people around the user, and the analysis of user's activity history by detecting Bluetooth devices in user's circumstance. This research is aiming to mine various kinds of smartphone logs including locations from GPS, pictures from a camera, application using histories, text messages from SNS (Twitter), and number of steps as exercise history from the wearable Jawbone device. Most people tend to spend time under certain patterns or rules based on day and time. But on holidays, people may visit somewhere new, or go out and hang out with some friends or more, or have a trip to far places. In our research an important step is to find features by analyzing the log's occurrences based on day and date. From the result of occurrence and feature analyses of multiple logs, we find out the user's daily life patterns, by which, we can further detect special events happened in these days with different life patters. These events would be valuable to support user's life recall.

Keywords—smartphone; SNS; applications; GPS; daily life pattern; feature analysis; event detection;

I.　まえがき

近年コンピュータの小型化，高性能化も加わり，その利用分野は実生活全般に大きく拡大し，必要不可欠なものになりつつある．1990 年初頭，Mark Weiser はコンピューティング環境の未来として，多数の，さらに多様な形態の情報機器群が強調し合い，人間の日常生活の中で意識せずコンピュータの恩恵を受けることのできる世界「ユビキタスコンピューティング(Ubiquitous Computing)」というビジョンを描いた．今ではユビキタスコンピューティングの世界がますます現実のものとなっている．電車，自動車，家電，高性能コンピュータを搭載した携帯端末といったように様々な状況・場所・物においてコンピュータやセンサが搭載されている．

さらに近年，非常に高性能で様々なセンサを搭載し，いつでもどこでもネットワークに接続が可能な携帯端末，スマートフォンの爆発的普及やウェアラブル機器の小型化には目を見張るものがある．スマートフォンとは，携帯電話とパソコン・PDA の機能を組み合わせた多機能な携帯端末である．また，ユーザがコンピュータを常に身に着けて利用するウェアラブルデバイスに対する注目も高まっている．多種多様なセンサが一つのシステムとして統合され，人が身に着けることにより，時間や場所に限定されることなしに各センサの計測データから健康管理，ユーザ個人の状態や行動記録を把握することが可能になっている[1]．このようなユーザの行動を記録したデジタルデータの履歴をライフログという．ユーザが見たものや聞いたもの，買ったもの，行った場所等ユーザの行動に関する記録全てを指す．ライフログの例として，Web ページの閲覧履歴や撮影した写真，メールの文章，購買履歴，移動履歴等が挙げられる．ライフログを調査することによる一番大きなメリットは，ユーザ個人の嗜好や状況に適したサービスの提供が可能となる点である．ユーザが認知していなかった趣味や嗜好の提示や行動の特徴を捉える行動分析や，行動の特徴から得られた結果から関連情報を提供するレコメンデーションサービス，日々の生活記録を呼び覚ますためのインターフェース提供やライフログ同士の関連性をユーザに示し，過去の行動を振り返ることを可能にする自己管理，可視化サービス等がある．

人は，曜日や時間などに従って規則的に行動する場合が多い．しかし，休日等の特別な日には普段と違った行動パターンをとるものだ．しかし，休日等の特別な日にはいつもは訪れない場所や遠方に旅行に行き，非日常的は体験を求めることもある．このような特別なイベントのあった日(以後イベント日)を検出することで，ユーザの振り返りの支援するサービスや娯楽情報の提供，推薦が可能なサービスを実現することができるだろう．

本研究では，人間の日常生活における規則性に着目しスマートフォンから得られる位置情報履歴，写真撮影履歴，アプリケーション使用履歴，および現在広く普及している SNS(ソーシャル・ネットワーキング・サービス)である Twitter におけるツイート投稿履歴，ウェアラブルデバイスである Jawbone 社の UP を使った歩数のライフログをそれぞれ長期間貯蓄し，一日のログデータ量の分析を行うことで，周期性のあるユーザの日常的行動特徴を割り出す．抽出された特徴から，イベント日を検出することを目的とする．なお，本研究においてユーザの行動特徴とは，長期的日数からみた曜日別のログ出現パターンの特徴と定義する．

本論文の構成は次のとおりである．第Ⅱ章では，関連研究について述べる．第Ⅲ章では，行動特徴抽出・イベント検出手法について述べる．第Ⅳ章では，ログの取得について述べる．第Ⅴ章では，行動特徴の抽出について述べる．第Ⅵ章では，イベ

ント日の検出について述べる．第VII章では，イベント日検出の評価について述べる．第VIII章では，まとめと今後の課題について述べる．

II. 関連研究

ライフログを利用し，ユーザの行動特徴を推定する研究は数多く行われている．位置情報ライフログを行動履歴とした研究は，様々な種類のログに位置情報を付加し，地図上に可視化するための汎用アプリケーションフレームワーク，MashMap フレームワークを提案する研究[2]や，複数の被験者に ActiveBadge を装着させ建物内のセンサで移動状態を検知し，行動履歴を日記に記すもの[3]，GPS で取得し蓄積した移動軌跡の分析を行うことで旅行時の車での行動パターンの抽出を行う研究[4]がある．また，ライフログの分析手法も様々あり，位置と曜日，時間帯の相関関係を指標にして分析を行い，日常の行動特徴を抽出する手法を提案している研究[5]もある．

こうした研究は，特定の場所，特定の日付，特定のデバイスの装着やアプリケーションに特化して，有効な移動軌跡や行動パターンの解析や手法を提案している．本研究では特定の条件に限定されることなく，現代広く普及していて非常に多くの機能を備えたスマートフォンや高性能化したウェアラブルデバイスをライフログ保存ツールとして使い，位置情報以外の様々なログを加えて分析することにより詳細な日常的なユーザの行動特徴を捉えることでその中に潜む非日常的なイベントをすることを目指した．

携帯電話端末，スマートフォンを用いてユーザの特徴を抽出する研究やサービスも多くあり，ActionLog という個人の携帯電話端末を利用し，位置情報や時間，周囲の人の情報を行動履歴として分析し，この行動履歴をコンテンツのドラフト記事としてユーザの記述した WebLog に自動的に情報を付加するシステムを提案する研究[6]や，NTT Docomo 社で行われているユーザの行動に合わせたサービスを実現するための行動推定技術の開発が行われている．NTT Docomo 社の研究では，携帯端末から得られる連続的な位置情報ログを基に日常的移動範囲の特徴を抽出し，行動推定を行っている[7]．また，「AirDiary：Bluetooth デバイス検出履歴を用いた半自動日記作成ツール」という研究では，生活の中でユーザの周囲で検出される Bluetooth デバイスに着目し，特徴抽出を行い，教師データを元にイベントの検出を行っている[8]．このように，スマートフォンを使ったライフログ分析の研究は広く行われているが，位置情報を要とした研究が非常に多く見受けられる．しかし，スマートフォンの GPS を継続的に起動させ，常に位置情報を連続的に取得し続けることはバッテリーの消費に大きな問題があり，長期間のライフログ取得に向いていないと考える．

本研究では 20 分毎に位置情報の取得を行い日常的行動エリアの滞在時間に着目することで電池消費の問題を軽減し，スマートフォンやウェアラブルデバイスから得られる日常的ライフログに加えて，写真の撮影や SNS 投稿といったイベント発生時に特に使用されると思われるライフログを統合的に分析し，生活に潜む周期性，特徴を発見することによる特異点，つまりイベント日の検出を行うことが目的である．

III. 行動特徴抽出・イベント検出手法

本項では，目的である非日常的なイベント日の検出までの概要を述べる．また，本研究で使用する基盤技術について詳細に述べる．

本研究は，ユーザが常に身に持っているスマートフォン，常に身に着けられるウェアラブルデバイスから得られるライフログに着目し，ユーザの日常的な行動特徴の抽出，イベント日の検出を行う．収集するライフログの種類は，イベントが発生する日に特に変化が見られると思われるものを選定した．スマートフォンから得られる GPS，写真，アプリケーション使用履歴，Twitter のツイート履歴，ウェアラブルデバイスから得られる歩数の 5 種類のライフログの出現頻度を，被験者 1 人を対象に日常的行動の周期性が得られるよう 5 週間に渡り貯蓄した．各ログの詳細な収集手法は第IV章にて述べる．得られたそれぞれのライフログの出現頻度からユーザの日常的行動特徴を抽出する．最初に，人間の日々の活動には周期性があるという前提のもと，ライフログを曜日別に分類する．分類されたデータからユーザの行動特徴パターンの抽出を行う．得られた行動特徴をもとに正規化ユークリッド平方距離を求め，非類似度の算出を行うことでイベント日を検出する．本研究によるイベント日の検出精度を評価するにあたって，被験者の実際にあったイベント日，本研究により推定されたイベント日の正誤判定を行い，F-measure による推定精度の評価を行った．図 1 にイベント日推定までのイメージ図を示す．

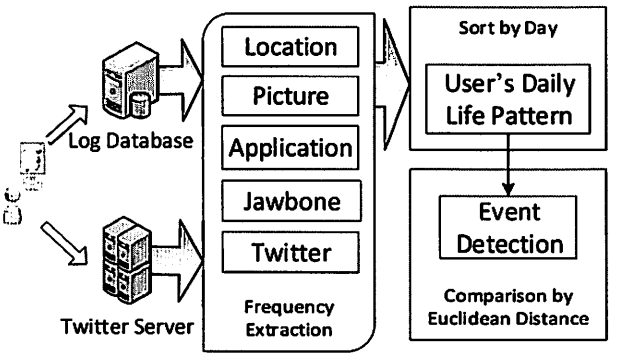


図 1．イベント検出法

本研究で使用した Android スマートフォンでのアプリケーション開発について述べる．代表的なスマートフォン OS である iPhone の iOS や Android は，ソフトウェア開発キット(SDK)が公開されており，開発者が自由にアプリケーションの開発・配布ができるようになっている．本研究では，開発環境を整えるのが容易かつ，開発の自由度が高く，日本での普及がトップである Android OS を用いてライフログ自動収集アプリケーションを開発した．Android アプリケーションを開発するには，Google 社が無償で公開，提供している Android OS を搭載したスマートフォンやタブレット端末で動作するプログラムを開発するために必要なソフトウェアをひとまとめにした拡張パッケージである Android SDK を Java に組み込むことで開発を行うことが可能となる．

本研究で分析を行う際に使用したデータマイニングツール RapidMiner について述べる．RapidMiner は株式会社 KSK アナリティクス（KSK Analytics Inc.）によって開発，提供されているオープンソースのデータマイニングツールである．データ統合，データ変換，データ分析，グラフ化等あらゆるデータマイニング手法を実現可能である．また視覚的に操作が容易であるという利点を持っている．本研究では，生データのライフログを比較可能なデータに変換するための正規化を行う際に使用した．

IV. ライフログの取得

ライフログを収集する上で大切なことは，手間がかからず無意識で残っているということである．自動でライフログを残すことを可能にすることにより，誰でも，忘れることなく負担のないライフログの収集ができる．

本研究で開発したスマートフォンアプリケーションでは，位置情報，写真，アプリケーションの 3 つのログデータの出現回数(起動回数)を自動で取得し，スマートフォン端末内のストレージに日付別に記録，保存する．また，日付毎のツイート回数を取得するため，Twitter 社のサーバに保管されているユーザのツイート情報を抽出し，日付ごとのツイート回数をテキストに出力する Java プログラムを開発した．最後に，ユーザの一日の運動状況の特徴を捉えるため，JAWBONE 社の 3 軸加速度センサ搭載のウェアラブルデバイス「UP」を使用し，運動を含めた一日の合計歩数を取得した．本セクションでは，今回使用した 5 つのログ取得についての詳細を 4 つの項に分けて述べる．

A. 写真・アプリケーション

ユーザの使用しているスマートフォンの写真撮影頻度，アプリケーション起動数を取得，保存が可能なアプリケーションを作成し，日付ごとの回数を集計した．また，将来的なサービスに役立つライフログとなるよう、撮影された写真には位置情報付加し，どのアプリケーションを多く利用したかも同時に記録するできるアプリケーションにした．

B. 位置情報 (GPS)

本研究では，新しい場所に訪れるといったログでイベントの検出を行うのではなく，ユーザの周期的に通う場所のログに着目し，イベント日における特異点を見つけ出し，イベント日を検出することを目的とする．そのため緯度・経度の示す数値情報よりも，その場所がユーザにとってどのような場所であるかという事が大事である(例:自宅，大学，会社)．緯度・経度の数値情報を開発したライフログ収集アプリケーション内で逆ジオコーディングを行うことによって町名，村名に変換，保存した．数値情報を住所として扱うことで分かりやすく，分析も容易なデータをライフログとした．逆ジオコーディングとは，緯度経度の点で表される数値情報を面で表現される空間情報に変換する技術のことである．また，住所の保存を行う頻度として，20 分間隔で保存をすることとした．一日中継続して GPS をアクティブにしてしまうとスマートフォンのバッテリーに大きな負担がかかる上，長期間の保存では情報量が多くなりすぎてしまい分析が困難となってしまうためである．また，人が特定の場所に 30 分以上いた場合、それを滞在したと見なすこととし，20 分間隔で住所の取得を行う．よって連続で同じ住所が 2 回以上出た場合は滞在とみなすことができる．また 20 分間隔ならばバッテリー消費を抑えられる利点が存在する．

ユーザの日常的行動の周期性を抽出するため，得られた住所から出現回数の最も多い 3 つの位置をユーザが毎週決まった周期で滞在する日常的行動エリアとして分析対象とした．例えば，自宅や勤務地，毎週通う習い事教室等である．住所は市や区で区切ってしまうと範囲が広くなりすぎて正確な推定ができなくなってしまうため町名，村名を保存することとした．

C. Twitter のツイート履歴

「Twitter」とは，2006 年 7 月にオブビアウス社(現 Twitter 社)が開発した SNS で，「ツイート」と呼ばれる 140 文字以内の短いメッセージをユーザが Web 上に投稿するシステムである．現在，世界に 1.4 億人のアクティブユーザが存在し，一日のツイート数は 3.4 億ツイート，内 55%がモバイルからの投稿であることがわかっている．自分の考えやあらゆるものに対する感想をツイートする他，旅行等のイベントがあった時にはその時々の感動を他のユーザと共有する，自分の思い出として残して置くといった使い方がされている．つまり，イベント日には日常生活と違ったログ出現特徴が現れるのではないかと予想し，一日のツイート数をライフログとして取得することと決めた．

ユーザが投稿を行ったツイートは Twitter 社のサーバに保管される．Twitter 社 API を使い，ユーザの一日のツイート数，ツイート内容を月や日付を指定して抽出，取得し，txt ファイルに出力するプログラムを作成した．取得したユーザの一日のツイート数を抽出し，日常的行動特徴，イベント日の検出に使用するライフログのひとつとする．プログラムより txt に出力された結果を図 2 に示す．

```
14 2013年10月31日頭、首、肩、腰、
15 ツイート数：6
16 2013年10月31日この張り紙を見た
17 ツイート数：7
18 2013年10月31日読了。寝る。。。
19 ツイート数：8
20 2013年10月31日RT @itoi_shiges
---
23 2013年10月22日のツイート数：5回,5
24 2013年10月23日のツイート数：13回,13
25 2013年10月24日のツイート数：9回,9
26 2013年10月25日のツイート数：8回,8
```

図 2. Twitter ツイート抽出プログラム出力結果

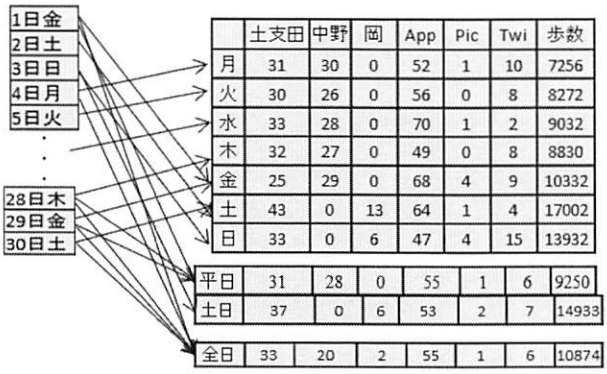
D. 歩数

スマートフォン搭載の加速度センサを使用しての歩数の観測も可能だが，長い期間かつ一日中スマートフォンを持っての観測は被験者及びスマートフォンのバッテリーに大きな負担がかかってしまうことや，激しい運動をする際にスマートフォンを携帯し計測することは非現実的である．そこで一日の歩数のライフログを収集するため，JAWBONE 社の UP を使用する．UP は，スマートフォンとの連携が可能な健康管理，活動記録用ウェアラブルデバイスである．また，UP 専用スマートフォンアプリケーションを利用することで計測したデータをいつでも同期，可視化することができる．また，ライフログを継続的に毎日，運動をしていても計れることが必要であるため，常に身に着けていられるスタイルと重量が大切である．UP はリストバンド型で重量が僅か 22g で防水という毎日身に着けることを想定して設計されている上一度の充電で約 1 週間の使用が可能である．そこで本研究では，長期間かつ水泳や激しい運動の最中であっても運動状態として歩数が収集可能なウェアラブルデバイス UP を採用した．また，UP をスマートフォンに繋ぎ，同期させることで蓄積されたライフログの視覚化を行ってくれる．また同期する際，蓄積された日付毎の歩数や運動状況，消費カロリーをスマートフォンが JAWBONE のサーバに転送，保存を行う．さらに、JAWBONE のホームページにアクセスし，登録を行ったアカウントページにてこれらのデータを csv 形式で得ることが可能である．本研究では UP から得られるデータのうち運動状態を知るため歩数を取得し分析を行う．図 3 取得された歩数データ，UP の外観を示す．

図 3. ウェアラブルデバイス UP，データ形式

V.　行動特徴抽出

ユーザのイベント日の検出することにおいて重要なことは，まずユーザの日常的行動の周期性を把握することである．また，本研究においてユーザの行動特徴とは，曜日別のログ出現パターンとする．ユーザの日常的行動特徴の検証を行うため，人間は曜日によって規則性を持って生活をするという前提のもと，蓄積された日付ごとの各ライフログをそれぞれの曜日，平日・土日，全ての日数(以後全日)において分類し，平均値を求めた．また，曜日ごとの出現特徴の可視化を行うことで，曜日とログの出現頻度の関連性について調べた．また，位置情報においては移動中や GPS の誤取得によって得られるような極端に出現数の少ないログは周期性のある日常的な行動特徴とは言えない．そこで，位置情報を 20 分間隔で取得を行い，2 回以上の連続出現回数を持つ住所を滞在場所とし，ユーザの行動特徴を現すライフログとして使用する．今回は 5 週間の中で滞在時間の長い上位 3 つの住所を周期性のある日常的行動として選定した．また，住所は町，村の範囲で取得を行った．収集したライフログを曜日，平日土日，全日に分類し平均値を求める様子のイメージ図を図 4 に示す．また，図 5 にログ属性ごとの平均的な曜日出現割合を表した円グラフを示す．



	土支田	中野	岡	App	Pic	Twi	歩数
月	31	30	0	52	1	10	7256
火	30	26	0	56	0	8	8272
水	33	28	0	70	1	2	9032
木	32	27	0	49	0	8	8830
金	25	29	0	68	4	9	10332
土	43	0	13	64	1	4	17002
日	33	0	6	47	4	15	13932
平日	31	28	0	55	1	6	9250
土日	37	0	6	53	2	7	14933
全日	33	20	2	55	1	6	10874

図 4. ライフログの曜日毎分類

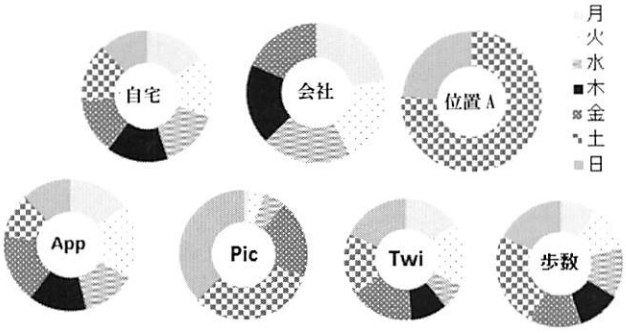


図 5. ログ毎の特徴円グラフ

図 4，図 5 の自宅，アプリケーション，Twitter，歩数のログはどの曜日でもほぼ均等に出現しているのに対してその他のログデータは出現しない曜日や極めて少なくなっている曜日も存在することがわかる．

また，住所「岡」は土曜日，日曜日のみにしか出現していなく，毎週末に決まって訪れる場所ということがわかる．また，被験者に調査を行ったところ，滞在時間の一番長い住所「土支田」は自宅であり，月曜日から金曜日の間長く滞在している「中野」は勤務地，土曜日日曜日に数時間滞在している住所「岡」は被験者が習慣的に趣味の運動を楽しみに訪れている場所であるということがわかった．そのため住所「岡」に訪れた日は歩数のログデータの値も大きくなっていることが分かる．

これらのことから，被験者は曜日や平日と土日によって周期的な特徴の中で行動していることわかった．

VI. イベント日の検出手法

イベント日の検出手法について述べる．日付ごとに蓄積されたライフログの正規化ユークリッド平方距離の合計値を求めることにより，特異点を検出，イベント日の推定を行った．なおイベント日の定義として本研究では，第Ⅴ章で得られたような，毎週決まった曜日に通うといったような行動はイベントではなく周期性のある日常的行動として定める．仲間とお酒を飲みに出掛ける，ショッピングにゆく等といった周期性のない事象をイベントとする．イベント日の検出をするにあたって，スケールの異なるログデータ同士を比較可能なデータ形式へ変換する必要があるためログデータの正規化を行う．また，被験者は曜日ごとや平日・土日に依存した生活を送っていることがⅤ章の特徴抽出によりわかった．そのため蓄積された全日数のログデータを合わせた推定に加え，曜日ごとに分類したログデータを利用した推定，平日・土日二つに分けた推定の 3 通りにおいてイベント日の検出を行う．これらより検出されたイベント日の推定精度を第Ⅶ章で評価することとする．本章では「ログデータの正規化」「正規化ユークリッド平方距離によるイベント日の検出」の 2 セクションに分けて詳細を述べる．

A. ログデータの正規化

収集されたログデータを分析する際，ログ属性同士の数値 1 単位あたりの重要度に大きな差があり，この値のままでは正しい比較や分析が困難である．例えば，月曜日の平均歩数である 7738 という値と月曜の平均ツイート数である 7 という値には明らかな差があり誤差程度にしか捉えられず，大きく偏った解析結果となってしまう．そのため，ログ属性間を同様の重み，スケールで比較，分析が可能なデータに変換するためデータの正規化(Normalization)行った．どのデータ要素もログ属性の平均値を基点にして，データはその前後に広がっていると言える．また，この偏差を評価する尺度を標準偏差(standard deviation)といい，この値を使って正規化を行うことができる．標準偏差とはデータ群の偏差を平均化したものであり，標準偏差を使って値を評価することによってそのデータが月並みであるか，または特殊であるかを判定することが可能となる．曜日 a のログ属性 $p(p_1 \sim p_7)$ の標準偏差 SD_{ap} は以下の式で求められる．

$$SD_{ap} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_{api} - \bar{x}_{api}\right)^2} \qquad (1)$$

この時，N は日数であり，曜日毎で計算を行う際は 5 週間分の同一曜日に対して分析を行うので，$N = 5$ となり，全日数で分析を行う際は 5 週間つまり 35 日なので $N = 35$ である．また，曜日 a のログ属性 p の i 番目のログデータの値をx_{api}とし，$\bar{x}_{api}$ は曜日 a のログ属性 p の N 個のログ属性の平均である．また，曜日 a は月曜日から日曜日に加え，平日・土日・全日をそれぞれ曜日として計算を行う．この標準偏差の値が大きければ大きいほど，ログ内でのデータの散らばりがあることを示している．従ってログ内でこの値が大きい曜日には日常と違ったイベントが起きている可能性があると予想することができる．曜日，平日・土日，全日数，それぞれの標準偏差を表 1 に示す．

表 1. 曜日毎標準偏差

	自宅	会社	位置 A	App	Pic	Twi	歩数
月	1.5	1.7	0.0	10.4	0.5	1.5	1089
火	1.8	6.5	0.0	8.7	1.0	1.0	1781
水	1.0	5.6	0.0	5.7	0.6	1.5	2581
木	1.3	1.5	0.0	3.3	0.0	2.5	834
金	4.8	3.9	0.0	14.3	1.9	1.4	3540
土	7.0	0.0	6.4	20.0	2.8	4.9	1001
日	21.5	0.0	6.0	14.9	2.6	4.4	6751
平日	3.1	4.5	0.0	10.4	1.4	2.7	1834
土日	13.3	0.0	6.1	16.4	2.0	4.6	5205
全日	7.5	10.5	4.1	10.7	1.8	3.1	3426

表 1 より，日曜日の位置ログデータ自宅や日曜日の歩数，土曜日のツイート等に高い値が見受けられる．これは，同じ曜日の日付間にログ出現数にばらつきがあることを表している。逆に，データのばらつきが小さいものは平均の近くに集まっているといえ，標準偏差の値が小さくなり，日常的な行動をしたと思われる．以上より求められた標準偏差を用いて蓄積された全日付のログデータを曜日ごと，平日・土日，全日の 3 パターンにおいて正規化を行う．日付 b のログ属性 $p(p_1 \sim p_7)$の正規化距離Z_{bp}は以下の式で求められる．

$$Z_{bp} = \frac{x_{bp} - \bar{x}_{ap}}{SD_{ap}} \tag{2}$$

曜日 a におけるログ属性 p の標準偏差 $SDap$ に対し，正規化する日付 b のログ属性 p のデータをx_{bp}，曜日 a におけるログ属性 p の平均 $\bar{x}_{ap}$ である．表 2 に正規化された 1 週間分のデータの例を示す．

表 2. 正規化データの例

	自宅	会社	位置 A	App	Pic	Twi	歩数
1 日(金)	−0.41	0.51	0.00	−0.03	−0.05	−1.21	0.18
2 日(土)	1.03	−1.44	2.47	−0.68	−0.24	−1.37	1.98
3 日(日)	0.27	−1.44	−0.40	−0.06	0.29	0.22	0.76
4 日(月)	−0.41	−4.41	0.00	−0.03	1.26	−0.08	1.18
5 日(火)	0.58	0.36	0.00	0.56	−0.71	−0.08	0.03
6 日(水)	−1.74	0.04	0.00	−0.85	−0.05	−1.59	1.14
7 日(木)	0.58	0.51	0.00	−0.03	−0.71	−0.83	0.37

正規化距離は平均値が 0，標準偏差が 1 となり，各データは同様の重みで平均値からの距離を表している．つまり平均値 0 から離れていればいるほど特殊なデータであるといえる．正規化されたデータは 0±1 の範囲に約 7 割のデータ，±2 の範囲内にほぼ全てのデータが入るという特性を持っている．多く写真を撮った日，多く運動した日，逆に普段頻出する位置 A の数が激減している，アプリケーションの使用回数が極端に少ないといった正負両方においてイベント日である可能性があるため，正負両方を考慮に入れた判断が必要である．

B. 正規化ユークリッド平方距離によるイベント日の検出

正規化されたデータの特性から，特定の日付において全ログ種の正規化データの平均値 0 からの距離の合計が大きければ大きい程，イベント日として推定することが可能である．しかし，正規化されたデータは正負に散らばっているため、データの距離を正負関係なく測る必要がある．これを式(2)の正規化距離に対して，正規化ユークリッド平方距離といい，曜日 a の 7 つのログ属性の正規化ユークリッド平方距離の合計Q_aを以下の式によって求めることができる．

$$Q_a = \sum_{p=1}^{7} \frac{(x_{bp} - \bar{x}_{ap})^2}{{SD_{ap}}^2} \tag{3}$$

この時，標準偏差SD_{ap}に対し，日付 b のログ属性 p のログをx_{bp}，曜日 a におけるログ属性 p の平均 $\bar{x}_{ap}$である．正規化されたデータは 0±1 の範囲に約 7 割のデータ，±2 の範囲内にほぼ全てのデータが入るという特性を利用し，特定の日付の 7 つのログ属性 p の正規化ユークリッド平方距離の合計が，$\sum_{p=1}^{7} {1_p}^2$ つまり 7 以上であればイベント日として妥当な値と判断し，選定を行った．例として，2013 年 11 月 6 日水曜日においての曜日ごと，平日・土日，全ての日の 3 パターンにより正規化されたデータを以下の表 3 に示す．なお，曜日毎は 5 つの水曜日，平日は 5 週間分の月曜日から金曜日，全日は 5 週間 35 日から比較した 11 月 6 日の正規化ユークリッド平方距離の合計Q_aをそれぞれ表している．

表 3. 2013/11/6 水曜日の正規化ユークリッド平方距離合計

ログ属性	水曜日	平日	全日
位置 A	−1.02	−1.74	−0.87
位置 B	0.18	0.04	0.59
位置 C	0.00	0.00	−0.40
App	−0.73	−0.85	−0.68
Pic	0.00	−0.05	−0.24
Twi	−0.76	−1.59	−1.37
歩数	0.83	1.14	0.04
Q_a	3.07	7.39	3.1

表 3 より，3 パターンそれぞれの合計に差があることが見受けられる．平日で比較を行った際のみ 6 日はイベントとして検出された．同様に全日付において 3 パターンの分析を行い，イベント日を推定する．

VII. 評価

本研究では 5 週間のログの取得を被験者 1 人に対して行い，収集されたログを元にユーザの曜日における行動特徴を抽出し，

全日，曜日ごと，平日・土日の 3 パターンにおいてイベント日の検出を行った．推定精度の評価を行うにあたり，被験者にログの収集を行ってもらう 5 週間の中で旅行やショッピング等の実際にあったイベント日を記録してもらった．被験者の実際に体験したイベント日と本研究によって推定された 3 パターンにおけるイベント日の正誤判定をそれぞれ行うことで推定精度を評価した．

本研究ではイベント日の検出精度を，統計的予測を行う上での評価用いられる F 値(F-measure)を用いて評価した．F 値とは，推定精度の評価指標のひとつで，結果の適合率と再現率という正確性と網羅性の総合的な評価を行うことができる評価法であり，F 値 F-measure は以下によって求められる．

$$\text{F-measure} = \frac{2 \cdot precision \cdot recall}{precision + recall} \quad (4)$$

この時，適合率 *precision* は推定されたイベント日のうち，実際のイベント日であり一致した割合，再現率 *recall* は実際にあったイベント日うち，イベント日と予測された割合である．

今回の実験により抽出された 3 パターンの抽出イベント日数，実際のイベント日数，それぞれの一致日数，適合率，再現率及び F 値により算出された推定精度の結果を表 4 に示す．

表 4. 推定精度

	全日	曜日	平日	土日
抽出イベント日数	8	6	9	4
実際イベント日数	12	12	7	5
一致日数	4	5	7	3
適合率	50%	83%	76%	
再現率	33%	41%	83%	
F 値	40%	55%	79%	

全日付で評価した精度は 40%と低くなり，ユーザ特徴に合わせた平日土日祝日の精度は 79%と高い精度が得られた．全日付でのイベント検出の精度が低くなった理由として考えられるのは，第Ⅴ章で得られたようにユーザは曜日平日・土日それぞれ別々の周期性で生活していることがわかっているため，ログ全体で評価した際，5 つある平日の周期性に対して 2 日間しかない土日が特殊なデータとして高い値を出してしまい，土日のみがイベントとして検出されてしまうといったことが起きてしまうためだと思われる．また，被験者は日曜日に実際にイベントが起きている事が多く，5 つの日曜日の内 4 つがイベント日であった．結果，他の要素と比較して類似度の低くイベントではなかった残りの日が少数派として逆にイベントとして検出されてしまったため曜日毎の推定精度が落ちてしまったと思われる．

平日間でイベント日検出を行った際，全てのイベントを取得できたものの，イベントでない日も 2 日検出されてしまった．これは Twitter 上で会話が盛り上がり，極端にツイート数が多くなる，なんらかの理由で会社を休み，一日家に滞在していたという本研究で指すイベントと異なる特異な日を検出してしまったためだと思われる．土日で比較した際は曜日別と同様に全体のイベント件数割合が高かったため，ユーザにとって特になにもなかった日が逆に特殊な日として検知されてしまうケースや，実際のイベント日であったにも関わらず他のイベント日の値と比べると値が低いために検知に失敗するケースがあった．しかし，ユーザの特徴抽出の周期に合わせ，平日・土日に分けて分析を行うことにより，より細かなイベント日を検出することに成功したといえる．

VIII. まとめ・今後の課題

スマートフォン，ウェアラブルデバイスから得られるライフログを分析することによりイベント日の検出を行った．また，ユーザ個人の特徴抽出を行うことでユーザ個人の特徴を配慮したイベント日の検出ができた．今後の課題として以下が挙げられる．

- 取得ログ種増加の検討．より詳細なユーザのライフログ情報を日常的に取得することにより，より正確な行動特徴，イベント日の抽出ができることが予想される．
- バッテリー消費の問題もあるが，一日をより細かく区切ることでより正確な移動履歴の推移を検出できる．バッテリーの問題を考慮しつつ，より高い頻度での位置情報取得ができることが望ましい．
- 今回，日々の活動の周期性に変化の生じにくい社会人を対象にライフロギングを行ったが，長期休みのある学生やフリーター等の生活があまり周期的ではないと思われる者を対象に実験を行い，本イベント検出手法の汎用性を証明する必要がある．
- さらに長い期間での実験の必要性．今回 5 週間に渡ってライフログの収集を行ったが，より長い期間のログを収集することでより正確な日常的行動特徴の抽出が可能であると思われ，イベント日の検出の制度も上がると思われる．
- イベント日検出の自動化．本研究のイベント日検出を自動化することでサービスへの応用が期待される．

文献

[1] 杉本千佳，長井宏和，江副亮介，佐々木健，板尾清，“ウェアラブルセンサによる行動認識システムの開発”，ニホン時計学会誌，Vol.51，No.197，2007.

[2] 高橋昂平，下條彰，本真佑，中村匡秀，“位置情報を含むライフログの可視化サービス開発支援フレームワーク”，電子情報通信学会，2011.

[3] W. M. Newman, M. A. Eldridge, M. G. Lamming, “PEPYS: Generating Autobiographies by Automatic Tracking”, Proceedings of the Second European Conference on Computer-Supported Cooperative Work, September 25-27, 1991.

[4] J. Wolf, R. Guensler, W. Bachman, “Elimination of the Travel Diary: An Experiment to Derive Trip Purpose From GPS Travel Data”, Notes from Transportation Research Board, The 80th Annual Meeting, 2001.

[5] 神崎優子，清水明宏，“ライフログサービスに適した位置情報分析手法の研究”，フロンティアプロジェクト学士学位論文，2010.

[6] 沼晃介，上松大輝，濱崎雅弘，大向一輝，武田英明，“ActionLog: 移動履歴に基づく位置情報付き Weblog の自動生成”，The 19th Annual Conference of the Japanese Society for Artificial Intelligence, 2005.

[7] 宮崎雄一郎，山田直治，住谷哲夫，磯田佳徳，“ユーザの行動にあわせたサービス実現のための行動推定技術の開発”，NTT Docomo テクニカルジャーナル，Vol.17，No.3，2009.

[8] 牛越達也，河野添之，“AirDiary: Bluetooth デバイス検出履歴を用いた半自動日記作成ツール”，情報処理学会研究報告，Vol.2011-HCI-142，No7，2011.

[9] 松本 光弘，清原 良三，沼尾 正行，栗原 聡，”位置情報を含む携帯端末利用履歴からのコンテキストに基づく最適アプリケーション推定法の提案”，情報処理学会研究報告，Vol.2010-MPS-80，No.3，2010.