

# ユーザの行動履歴の自動取得に基づく生活の記録化

理学専攻・情報科学コース 清水 美絵 (指導教員：小林 一郎)

## 1 はじめに

近年, SNS などの普及により自分の生活経験をブログなどに書き記す機会が増えた. 本研究では GPS を利用して移動履歴を, Google Calendar から登録された予定・イベントを取得し, 生活のログを半自動で記録することを試みる. また, 日ごとのログを, 自分の活動において頻繁に利用する場所や所属するコミュニティ, イベントで分類し月単位で記録を遡って見ることができる, 活動分類カレンダーを作成する取り組みについて説明する.

## 2 活動分類カレンダー作成の準備

活動分類カレンダーの作成過程を図 1 に示す.

GPS 受信機, Google Calendar からそれぞれ地理情報と予定・イベント情報を取得し, システムが編集を行い, 活動分類カレンダーとしてユーザに提示する.

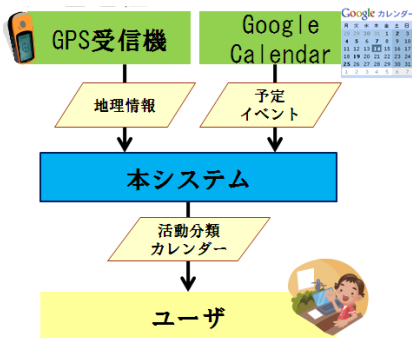


図 1: 活動分類カレンダー作成過程

### 2.1 GPS を利用した地理情報の取得

本研究では, 携帯電話に付属された GPS 受信機のように持ち歩くだけで移動履歴を取得できる Sony 製の GPS-CS1K を用いる. GPS-CS1K は 15 秒ごとに位置を観測し, 観測時間と位置をログファイルに蓄積する仕様となっている.

### 2.2 Todo リストの準備

観測された場所での活動を推測するために, 事前に Todo リストを準備する. Todo リストはいつ (when), だれと (who), なにを (what), なにのために (why), どのようにしたか (how) の 4W 1H の形式で登録する.

例えば地理情報が『テニスコート』, リストが『テニスの試合をする』の場合, 地理情報ではテニスコートにいたことしか判断できないが Todo リストの情報を与えることでテニスコートでの活動がテニスの試合であることが分かる.

### 2.3 個人プロフィール情報の作成

ユーザの活動特性をシステムで把握するために, 事前に個人プロフィール情報を準備する. 個人プロフィール情報は, 活動分類カレンダーを作成する際に活動を分類する指標となる情報が含まれており, 必ず作成する必要がある. 本研究では, 所属するコミュニティごとに活動の分類を行う. コミュニティごとに主な活動を行う場所があると考え, 個人プロフィール情報へ活動場所のランドマーク名を登録する. また, 主な活動場所はないがコミュニティとしてのつながりを持つ場合もコミュニティ名を登録する.

```
<name>清水美絵</name>↓
<community>石川中学校, 凸版印刷</community>↓
<latitude>35.687688</latitude>↓
<longitude>139.359848</longitude>↓
<name>スーパーマーケット</name>↓
<attribute>買い物, 食品, 文房具, 雑貨</attribute>
```

図 2: 個人プロフィール情報

### 2.4 Google Calendar の設定と予定の登録

Google Calendar はユーザの用途に合わせて複数のカレンダーを持つことができる. この機能を利用して GPS から取得する地理情報をユーザの主な活動場所ごとに自動で分類し管理する. もし, 主な活動場所のどれにも適さない場合は, 親カレンダー (ユーザ名のカレンダー) に登録される. カレンダーで分類を行う準備として, 個人プロフィール情報に登録されているランドマークごとに Google Calendar へカレンダーを設置する.

### 2.5 共起語辞書の作成

GPS から取得する地理情報の日記群と Google Calendar から取得したイベント群との共起関係を把握することでユーザの行動傾向を把握する. 適切な共起関係を規定するためにユーザごとにイベントの共起語辞書を作成する. 共起語の抽出には Google 検索でイベント名を検索したときに出現する上位 10 ページに現れる全ての単語を共起語候補群としそれぞれの  $tf \cdot idf$  値を計算する.  $tf \cdot idf$  値の高い上位 50 語を最終的な共起語の候補とする.  $tf$  値はイベント名で検索した Web 上位 10 ページに現れる共起語候補群全てに対し, 出現回数をカウントする.  $idf$  値は Google に掲載された全ページ 2,080,000,000 件のうち共起語候補群すべてに対して Google 検索のヒット件数を求め, (全ページ数 / ヒット件数) の対数をとって計算する.

ユーザは最終的な共起語候補の 50 語の中から自分の特性にあう共起語を選択する. 候補に出現しない単語を登録したい場合は, Web 上で単語を入力することで辞書に登録できる. 共起語は検索効率を考慮し, 各イベントに対し 10 個までとする.

イベント	共起語
スキー	スキー 競技 五輪 技術 アルペン スキー シーズン スポーツ 情報
旅行	ツアー ホテル 海外 観光 プラン 予約 空港 出発 準備 英語
パーティー	家族 プレゼント ケーキ パーティー
学会発表	原稿 聴衆 論文 スライド プレゼン 研究 説明 質問
卒業式	証書 学校 衣装 謝恩会

図 3: 共起語辞書

## 3 地理情報の行動履歴日記への変換

### 3.1 移動距離, 速度の計算

地理情報として蓄積された移動履歴に対して移動前と移動後の緯度経度の差から移動距離と移動速度の計算を行う. 移動距離は地球を球と見立てて計算を行う球面三角法を用いる.

### 3.2 住所またはランドマークの取得

本研究では Google Maps から提供されている API を用いて緯度経度からその位置を示す住所を見つける逆ジオコーディングを行う. 特定できる住所の詳細度は番地とする. 住所が「東京都文京区大塚 2 - 1 - 1」の場合, 逆ジオコーディングで特定する住所は「東京都文京区大塚」までとなる. もし, 観測した位置が個人プロフィール情報に登録した主な活動場所のランドマークの緯度経度とほぼ同一の場所 (誤差範囲は 400m) を示した場合は, その位置を示す情報を住所ではなくランドマークとする.

### 3.3 データのフィルタリング

地理情報は 15 秒間隔ごとに取得されるため, 1 時間の観測データでもデータ量が膨大になり, 処理や計算に時間がかかってしまう. そこで次の指標を基にデータのフィルタリングを行い, 大量データの中からユーザの移動を把握できる重要部を抽出する.

- (1) データ抽出を 60 秒間隔にする
- (2) 逆ジオコーディングの結果, 同一の住所を示す場合は抽出しない
- (3) 移動速度が一定 (10km 以上変化しない) の場合はユーザが連続して同じ行動をしているとし, 抽出しない

### 3.4 Todo リストの反映

Todo リストに記載された内容が観測されたランドマークで行う可能性の高い場合、その場所での行動を Todo と推定する。本研究では Todo の中に含まれる単語群の上位概念を WordNet で検索し、Todo の結び付けキーワードとして利用する。登録されたランドマークで行う主な活動情報を個人プロフィール情報から取得し、Todo キーワードとの一致を判断し、ランドマークと Todo の適切な組み合わせを見つける。これにより、Todo「牛乳を買う」、ランドマーク「スーパーマーケット」の場合、牛乳の上位概念として抽出された「食品」、ランドマークに付加された「食品」とを一致させ、スーパーマーケットでの行動を「牛乳を買った」と推定することができる。

### 3.5 行動履歴日記の作成

上述した処理を行い、行動履歴日記を作成する。行動履歴日記は [1] 移動前時間、[2] 移動後時間、[3] 移動前住所、[4] 移動後住所、[5] 移動距離、[6] 移動速度を基に、[3] 移動前住所、[4] 移動後住所、[6] 移動速度ごとにフォーマットを作成し、それに情報を当てはめることで日記の編集を行う。

<フォーマット例>

1. 移動前と後の住所が異なる  
『[1] 移動前時間から [2] 移動後時間まで [3] 移動前住所から [4] 移動後住所を時速 [6] 移動速度 km で移動した』
2. 移動前と後で住所が同一で、速度が 6km 以上  
『[1] 移動前時間から [2] 移動後時間まで [3] 移動前住所あたりを時速 [6] 移動速度 km で移動した』

### 3.6 ユーザによる日記の編集

作成した行動履歴日記に対し、ユーザが不要だと思う文章や、足りないと感じる文章はユーザの判断で削除追加を行う。

### 3.7 Google Calendar への行動履歴日記の登録

Google Calendar にアクセスし、作成した行動履歴日記が適切なカレンダーに登録されるように場所データを基に判断し、自動で登録される。

## 4 活動分類カレンダーの作成

地理情報、予定・イベント情報を取得後、月単位で Google Calendar に登録された活動をコミュニティやイベントの関連性を考慮して分類を行う。活動の分類には三つの指標を用いる。一つ目は、個人プロフィール情報に登録してあるユーザが頻繁に活動を行う場所に基づく分類（指標 1）。二つ目は主な活動場所はないがコミュニティでつながりがある場合の分類（指標 2）。三つ目は Google Calendar に載せた地理情報、予定・イベント情報すべてに対し、キーワードを抽出しそのキーワードを指標としての分類（指標 3）。これら三つの指標を基にユーザが自分の行動傾向を知るための活動分類カレンダーの作成を行う。

### 4.1 活動分類カレンダーの作成過程

まず指標 1 のランドマーク名で分類し、さらに指標 2 のコミュニティでの分類を行う。この処理を行うことでコミュニティごとに行動履歴やイベントが集約できる。この集約された行動履歴やイベントの中でさらに関連性の高いもの発見し、色分けを行う。関連性の判断には、行動履歴やイベントごとにキーワードを抽出して行う。キーワードは行動履歴やイベントに含まれる単語をキーワード候補群として抽出し、キーワード候補ごとに  $tf \cdot idf$  値を計算し、一番値が高いものをキーワードとする。さらにキーワードとして出力された単語が事前にユーザの作成した共起語辞書に登録されたイベントと同一の場合はその共起語を抽出する。

コミュニティごとに集約された行動履歴やイベントから上述した手法で抽出したキーワード、共起語を基に共起関係を見つける。共起関係の強さは共起ポイントを加算することで取得し、値が高いほど関係が強いとする。共起ポイントの計算手法は、自分のキーワードが他の行動履歴やイベントに含まれる場合、共起関係が高いとし共起ポイントとして“1”を加算する。また、自分の共起語が他の行動履

歴やイベントに含まれる場合、共起関係がやや高いとし共起ポイントに“0.5”加算する。行動履歴やイベント同士の共起ポイントが合計で“1”ポイントを超えた場合は共起関係が強いとして同一の色で色分けされる。

## 5 実験

2009 年 10 月の 1 か月と 2009 年 11 月の 1 か月の活動分類カレンダーを作成した。図 4 は、2009 年 11 月において処理をした作成されたカレンダーの一部である。共起関係が強い行動履歴やイベントは同一の色で色分けされていることが分かる。2009/11/8 と 2009/11/22 は「コストコ」という巨大スーパーマーケットへ行ったため、「コストコ」というキーワードで関連づけられている。また 2009/11/18 と 2009/11/20 はそれぞれ「小宮」という地名をキーワードとして持つため共起関係が高いと判断された。

日付	日記
2009/11/08	私は17:12:08から17:34:50の時間、東京都 新宿区 北新宿から東京都 新宿区 歌舞伎町を時速3Kmで移動した。
2009/11/08	私は19:03:02から19:03:17の時間、スーパーアルプスから東京都 八王子市 宇津木町を時速6Kmで移動した。
2009/11/08	コストコへ行く。
2009/11/18	私は10:59:09から11:05:34の時間、小宮駅にいた。
2009/11/20	私は21:17:55から21:19:00の時間、小宮駅から東京都 八王子市 小宮町を時速39Kmで移動した。
2009/11/22	私は09:33:05から09:39:35の時間、東京都 町田市 小山ヶ丘のあたりを時速27Kmで移動した。コストコへ行く。

図 4: 2009 年 11 月の行動分類カレンダー

図 5 ではマイカレンダーが親カレンダー（清水美絵）になっており活動場所での共起関係はないが石川中学校（母校）の友人コミュニティとして結びつきを発見し共起関係が強いと判断され色分けされた。

日付	開始時刻	終了時刻	マイカレンダー	コミュニティ
2009/11/04	00:00:00	00:00:00	清水美絵	石川中学校
2009/11/20	18:17:36	18:26:36	清水美絵	石川中学校

図 5: コミュニティに基づく共起関係

図 6 では、2009/10/14 のイベント“ケーキを手作り”のキーワードは「ケーキ」、2009/10/15 のイベント“母の誕生日”のキーワードは「パースデー」で同一の単語を含まないため、このままでは共起関係を見つけられない。しかし、「パースデー」というイベントを共起語辞書に登録していたため“パースデー”の共起語に“ケーキ”を発見し、この 2 つのイベントの共起関係を見つけることができる。

日付	日記	キーワード
2009/10/14	ケーキを手作り	ケーキ
2009/10/15	母のバースデー	バースデー

図 6: イベントに基づく共起関係

## 6 おわりに

本研究では、地理情報と予定・イベント情報からユーザの行動履歴を行動分類カレンダーとして提供するシステムの提案・構築を行った。さらに、行動履歴をユーザの主な活動場所やユーザの所属していたコミュニティ、その月に起こったイベントで分類することにより、ユーザの行動分類を行った。それにより、ユーザが 1 か月という長いスパンで自分の行動を把握し、行く頻度の高い場所やその月に起こったイベントに共起される行動を見つけることでユーザの行動の傾向を示すことができた。

今後は、地理情報から移動履歴を取得する際に行うデータフィルタリング精度をさらに向上し、不要な地理情報を削除して検索時間を速くし、またユーザの行動分類指標を現在の場所、コミュニティ、イベントの三つからさらに指標の追加を検討し、行動分類の向上を目指したい。

## 参考文献

- [1] 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 の自動作成, 第 19 回人工知能学会, 1C2-04, 2005.
- [2] 沼 晃介, 平田 敏之, 武田 英明, 学術会議における位置情報コンテキストを用いた Weblog 作成支援システムの開発と運用, 人工知能学会研究会資料, SIG-SWO-A501-01, 2005