

地域イベント発見および特性検証のための実空間マイクロブログを用いたユーザ移動パターン分析システム

藤坂 達也 李 龍 角谷 和俊

兵庫県立大学環境人間学部

1. はじめに

近年, Twitter [6]に代表されるマイクロブログが普及したことで, 短い文でユーザの状況を即時に共有することが可能となった. さらに, iPhone のようなスマートフォンから行われる場合, GPS 機能を用いた正確な位置情報をメッセージに自動で付与して発信することができるため, ジオタグ付きのメッセージが実空間に多く残されている. そこで我々は, そのような実空間マイクロブログを発信するユーザを 1 つセンサーと捉え, マイクロブログサイトが実空間を観察する Human Sensor Network であると定義する. これより, 世界で起こる様々なイベントや出来事をマイクロブログを通じた人間からのセンシングにより理解することが可能である.

そこで, 本研究ではこの考えに基づく社会的な分析を行うために, 実空間マイクロブログを用いてユーザの移動パターンを解析する分析システムの開発を行う. さらに, 我々のシステムを用いて地域イベントを発見しその特性を検証する.

2. 関連研究

マイクロブログはまだ発展途上のサービスであり, 様々な調査が行われている. Java ら[3]や Krishnamurthy ら[4]は, Twitter の利用地域や時間におけるユーザら発言頻度や典型的な話題について調査している. また, 岩木ら[2]はマイクロブログから有用な記事の発見を行っている. これらの研究は主にメッセージの中身やリンク構造に基づく知識発見がなされているが, 我々はマイクロブログでも位置や時間に着目し, 特にジオタグされたマイクロブログから地域的な社会の分析を行うことで大きく異なる. 一方, 守屋ら[5] はブログのような地域情報の書かれやすいテキストから地域の場所の特徴を推定し, その結果をデジタルマップ上に表示している. この研究は, 社会的分析という点で類似しているが, 我々はマイクロブログによる実際のユーザらの移動パターンを分析するという点で異なる.

3. ユーザ移動パターン分析システム

我々は, ユーザらの移動パターンを分析するに当たり, マイクロブログサイトから大量の実空間マイクロブログを取得するシステムおよび実際の移動パターンを分析するためのシステムを開発した.

3.1. 実空間マイクロブログ取得システム

図 1 にシステムの概要を示す. このシステムでは, 国のような大きな地域から市や町のような狭い地域まで, ユーザの地域範囲の粒度に柔軟に対応できるようにする

ため, インプットとしてオンライン地図上で操作し取得範囲を指定する.

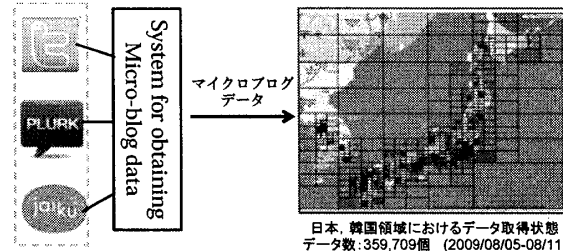


図 1 マイクロブログ取得システム

そして, ユーザが指定した領域からその中心 (緯度, 経度) と半径をシステムで求め, Twitter API[7]に問い合わせることにより, その領域に属するデータを取得する. しかし, この API には制限があり, 1 回の問い合わせで最高 1,500 件の最新のデータしか取得することができない. つまり, 指定した領域が 1,500 件以上のデータが存在する場合, それ以上のデータの取得ができない.

そこで我々は, 図 2 に示した地域依存のマイクロブログを網羅的に取得する手法でデータの取得を行う. この手法では, もし指定した領域のデータ数が 1,500 件に達した場合, システムがその領域を図 2 のように 4 分割し, それぞれの領域で中心と半径を指定し直し, それぞれの領域に関して問い合わせを行う.

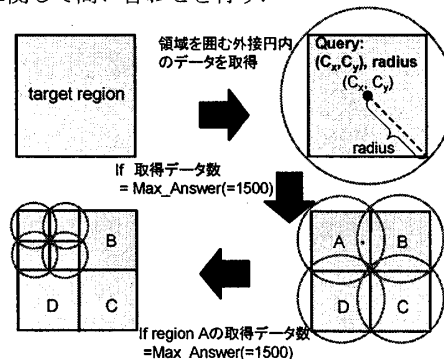


図 2 データ発生分布に適応したデータ取得手法

これを再帰的に繰り返すことで, 都市部のような多くのデータが存在する地域や郊外のような少しのデータしか存在しない地域でもデータに合わせて分割するため効率よく自動でデータを取得することが可能となり, 膨大なデータの取得を実現できる.

3.2 実空間マイクロブログ分析システム

我々は, 地域イベントの発見ために 2 つの移動パターンモデル, 集中・分散型モデル[1]を提案した. 集中型モデルの概要について説明する. このモデルは, 人々が移動することによりある時間に集中する場所が発生することを表すモデルである. 我々は, もし人々がある時間に

Analytical System based on User Movement Patterns using Mass Geo-Tagged Micro-blogs for Discovering Social Event and Exploring Characteristics

Tatsuya Fujisaka, Ryong Lee and Kazutoshi Sumiya

School of Human Science and Environment, University of Hwogo

特定の場所に集まるとき、そこではお祭りや通勤ラッシュなどの地域的なイベントが発生している可能性があると考えた。次に、分散型モデルの概要について説明する。このモデルは、人々が移動することによりある時間に人々の分散が発生することを表すモデルである。我々は、もしある時間に普段の活動地域から分散する人が多い場合、そこでは自然災害やお盆のような帰省ラッシュなどの地域的なイベントが発生している可能性があると考えた。

我々は、この 2 つの移動パターンモデルに基づき地域イベントが発生していると考えられる場所の発見するため、図 3 に示す実空間分析システムを開発した。このシステムの概要について説明する。

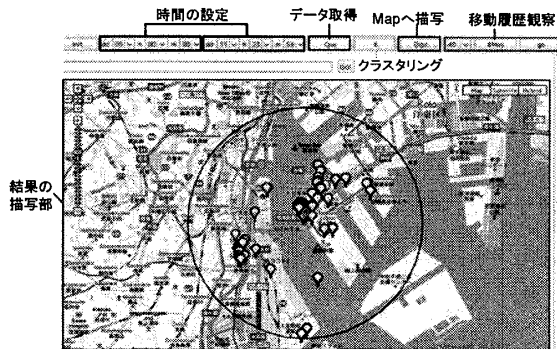


図 3 実空間マイクロブログ分析システム

まず、ユーザは取得したデータから分析したい時間を設定することで、ジオタグ付きのマイクロブログを取得する。次に、抽出したデータの位置に基づく分布傾向を把握するために、クラスタリングを行い、オンラインマップ上に描写する。今回は、クラスタリング手法として K-means 法を用いる。実際の、日本のクラスタの分布傾向を図 4 に示す。

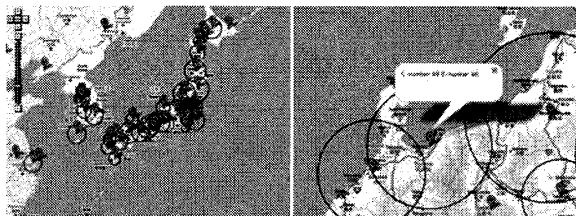


図 4 クラスタ傾向 (K=100)

クラスタは黒い丸、その中心を赤いアイコンで表す。さらに、アイコンにはクラスタの id、とクラスタのメッセージ数をタグとして付与している。

次に、提案した移動パターンモデルに基づく指標として下記に示す各クラスタの活動度(Activity)を定義する。

$$Activity_i = \frac{\sum_{j=1}^{i_{users}} moving\ distance\ (user_j)}{radius_i\ (km) \times \#users_i} \quad (1)$$

$\{radius_i, : the\ radius\ of\ i-th\ cluster$
 $\#users_i, : the\ number\ of\ found\ users\ in\ i-th\ cluster$

活動度では、クラスタごとに存在するメッセージから発信ユーザを特定し、ユーザらの移動履歴から移動距離をそれぞれのユーザで算出し、上記の式(1)に基づき、地域

的イベントが発生していると考えられる人々の移動が激しい場所を見つける。次に、我々はメッセージ解析により、活動地域の激しい中でも地域的なイベントが起こっている場所を探す。ここでは、クラスタ内の出現頻度を tf とし、他のクラスタでの出現頻度を df とし $tf-idf$ を求め、特徴語を抽出し、実際のメッセージを見ることでイベントの発見を行う。

さらに、我々は地域イベント行われていたクラスタ内のユーザらの実際に移動履歴を図 5 のようにオンライン地図上に描写することで、そのイベントの特性、例えばどんな地域からどれくらいの人が集まるかなどを検証できると考える。ここでは、赤いアイコンと線でクラスタ内のユーザらの移動軌跡を表している。

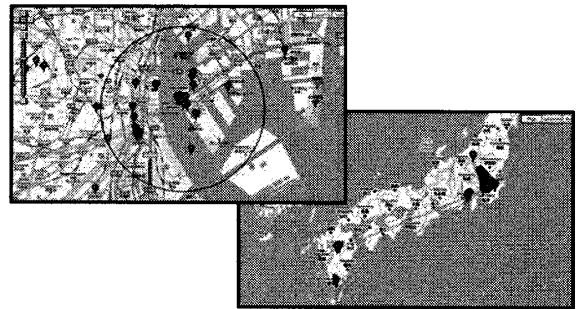


図 5 お台場クラスタの実際の移動軌跡

4. おわりに

我々は、Human Sensor Network という考えた方に基づき、大量の実空間マイクロブログから地域イベントを発見し、その特性を検証するためのユーザ移動パターン分析システムを開発した。今後の計画として、システムを用いた様々な地域イベントを発見し、評価を行う。

謝辞

本研究の一部は、平成 21 年度科研費基盤研究(B)(2)「ユーザの潜在的意図を用いたレス・コンシャス情報検索基盤の構築」(課題番号: 20300039) によるものです。ここに記して謝意を表すものとします。

参考文献

- [1] Tatsuya Fujisaka, Ryong Lee and Kazutoshi Sumiya, Discovery of User Behavior Patterns from Geo-tagged Micro-blogs 4th International Conference on Ubiquitous Information Management and Communication (ICUIMC 2010), January 2010
- [2] 岩木 祐輔, アダム ヤトフト, 田中 克己: マイクロブログにおける有用な記事の発見支援, DEIM Forum 2009 A6-6, (2009)
- [3] Akshay Java, Xiaodan Song, Tim Finin, Belle Tseng. Why we twitter: understanding micro-blogging usage and communities, Proceedings of the 9th WebKDD and 1st SNAKDD 2007 workshop on Web mining and social network analysis, San Jose, California, pp 5665, (2007).
- [4] Balachander Krishnamurthy, Phillipa Gill, Martin Arlitt. A few chirps about twitter. WOSP '08: Proceedings of the first workshop on Online social networks, pp19-24, (2008)
- [5] 守屋敬太, 佐々木史織, 清水康: 地域情報関連テキストを対象とした地域状況表示地図の動的生成方式, DEIM Forum 2009, (2009)
- [6] Twitter: <http://twitter.com/>
- [7] Twitter Open API: <http://apiwiki.twitter.com/>