

## 審査の結果の要旨

氏 名 ウィタヤンクーン アピション

2007年以降、携帯電話による緊急通報においてGPSによる位置情報の通知が義務化され、多くの携帯電話がGPS受信チップを内蔵し、リアルタイムで位置データを取得できるようになった。その結果、緊急通報以外でも携帯電話によるリアルタイム位置情報を利用したサービスが拡大した。たとえば、ナビゲーションサービスや位置に連動したレストランなどの検索サービス、位置を利用したゲームなどである。数多くの利用者に関する位置情報がサービスの利用ログとして蓄積され、その利用が大きな課題となっている。

位置情報のログは匿名化されているため、位置座標と時刻情報しか含まれていないが、長期間にわたるログを視覚化すると、どの地域に居住しどこに勤務し、日常的にどの範囲を移動しているかなど、生活のパターンをある程度読み取ることができることがわかる。こうした生活パターンや重要な生活地域（住居や勤務場所等）の推定は、マーケティングなどの産業的な利用だけでなく、交通機関や施設等の社会インフラのマネジメントの改善、都市開発計画の支援、新型インフルエンザなどの伝染病の拡散予測や抑制方法の検討、大規模災害時の帰宅難民等の誘導など社会公益的な目的にも大いに貢献すると期待される。

しかし、位置情報のログは100万人規模、1年以上の長期間にわたるいわゆるビッグデータであることから、解析手法の開発だけでなく、大量データの処理や管理、検索の高速化、高効率化手法の開発も重要である。大量データ処理・管理等の高速化手法についてはHadoopのようなオープンソースのシステムが公開されるなど開発が進んできている。しかし位置情報を対象とした検索の高速化などについてはHive（Hadoop上で動作するデータ検索や集計等のデータウェアハウスシステム）などでもサポートされていない。一方、解析手法についても携帯電話等から大量に蓄積できる位置データは、計測間隔が5分程度と比較的長く、従来から収集されてきた少量の実験的データが数秒から1分程度の短い計測間隔を有しているのと比較して、正確な情報抽出（例えば、移動交通モードの推定など）には多くの工夫が必要になる。このように、大量の位置データを処理するプラットフォームの開発と、その上で動作する解析手法とシステムの研究開発が必要とされている。

本論文はHadoop/Hiveを改良し、携帯電話に装着されたGPSから得られる大量の位置データの効率的な検索や処理を可能とした上で、匿名化された個別の利用者の位置ログデータから、利用者の居住地域や勤務地域、移動の状況などを推定する手法、さらに花火大会などのイベント等をGPSログデータから発見する手法を開発し、約100万人の1年間にわたる位置ログデータに実際に適用することでその有効性を実証しようというものである。本論文は9章からなっている。

第1章はイントロダクションであり、研究の背景や目的、成果の独自性、論文の全体構

成などをまとめている。

第 2 章は位置データ解析のためのデータ管理システムであり、Hadoop を中心に大量データの管理方法について既往の研究を整理し、その上で本論文が対象としている 100 万人データの処理を念頭にシステムのパフォーマンステストやチューニングを行っている。Hadoop はオープンソースソフトウェアであり、多くの実装例があるため、研究開発成果の共有を図る上で有効であることから、本論文でもシステムの基盤として Hadoop を採用している。

第 3 章は空間データの処理であり、Hadoop 上で動作するデータウェアハウスシステムである Hive を対象に、移動を中心とした位置データの検索を高速で行えるような改良を行った。これにより、大量の位置データを高速に処理・解析できる基盤が構築された。

第 4 章は移動を中心とした位置ログデータの解析手法であり、位置ログデータ解析のためのライブラリーを開発・実装している。

第 5 章は生活パターンの抽出であり、携帯電話に装着された GPS からの位置データが、利用者が移動等の活動を行っている時のみに取得されることに着目して、24 時間の中でどの時間帯に活動を行っているのかという生活パターンを抽出している。その結果、全体としては 24 のパターンに分類されるものの 7 パターンで全体の 9 割弱を説明できること、夜間のみ積極的に活動する人々も全体の数パーセント存在し、居住地の抽出などにおいては注意を要することなどが示された。

第 6 章は居住地域や勤務地域等の重要な生活地域情報の抽出である。位置データの時間的・空間的クラスタとして滞在点を抽出し、滞在時間や時刻を加味した上で、学習データを元に機械学習により居住地等のラベルをつけることで、居住地域等の推定を行った。その際、ウェブを使った活動日誌データと GPS からの位置ログデータを突き合わせた位置付き活動ログデータを収集し、さらに位置ログデータを判読してマニュアル作業でラベリングした「準検証データ」と併せて、学習データあるいは検証データとした。その結果、精度は約 90%になることが示された。

第 7 章は、移動における交通モード推定である。ここでは連続的に変化する位置ログデータを滞在点や交通結節地点で分割して、移動の距離や速度等を手がかりに機械学習の手法により交通モード（徒歩、自転車、自動車、鉄道）を判別した。その結果、自転車のみが 40%~60%程度の精度となったが他の交通モードについては 83%から 100%の精度となることが示された。

第 8 章はイベント等の異状検知である。ここでは 500m メッシュ単位に 30 分ごとの滞在人数を位置ログデータから集計し、その変化を HMM（隠れマルコフ連鎖モデル）でモデル化することで、メッシュ毎の滞在人数変化をモデル化した。そして位置ログデータから得られた滞在人数の生起確率の変化の大きさとその継続時間で「異状度」を評価した。関東地域における花火などのイベント事例を元にパラメータの調整を行い、96%~98%の精度を得た。

第9章は結論であり、得られた成果と今後の課題をまとめている。

以上をまとめると本論文は大量に携帯電話から得られた GPS による位置ログデータを処理・検索する基盤システムを初めて開発し、さらに既存の機械学習手法を組み合わせながら、外れ値の処理等を工夫することで、大量の位置ログデータ（100 万人・1 年間）を実際に解析し重要な生活地域等を明らかにしており、空間情報工学や人の行動解析分野に多大の貢献をしている。よって本論文は博士（工学）の学位請求論文として合格と認められる。