論文の内容の要旨

［論文題目］

大規模位置データ分析・流通手法に関する研究

氏　名　　加美　伸治

センサーやスマートフォンから収集蓄積される大規模ストリームデータを，社会システムなど様々なアプリケーションに活用するための研究がすすめられている．本論文では，その中でも位置情報の付帯したデータ（以後，位置データと呼ぶ）に着目する．マイクロブログなどのソーシャルデータ，センサーデータなどのマシーンデータなどから得られる位置データを活用するサービス・アプリケーションは，テレマティクス，マーケティング，位置情報ゲームなど幅広い分野で商用化されており，そのための大規模位置データ管理基盤に関する研究が進められている．そのような大規模位置データ管理基盤は，大きく，データをセンシングし蓄積する情報収集・蓄積ブロック，収集したデータから重要な情報を抽出し，アプリケーションに流通させる情報分析・流通ブロック，そして取得した情報の可視化などの提供や，様々なアプリケーションへの活用を行う情報提供・活用ブロックで構成され，それぞれにおいて様々な研究課題がある．

本論文は，大規模位置データデータ管理基盤を実現する上で必要となる情報分析手法および情報流通手法について論じたものである．情報分析手法は，収集・蓄積したデータから，有用となる情報を抽出するための手法であり，大規模なデータに潜むアプリケーションが必要とする情報をいかに効率的に抽出するかが技術の鍵である．情報流通手法は，そのように抽出した位置情報が付帯するデータ・情報を，それを必要とする実空間上に分布するアプリケーション・ユーザに効率的に伝播配信するための手法であり，実空間性に起因する伝播特性から情報受信者の情報ニーズに合致する伝播パターンを実現することが目的である．

本論文は以下の全 5 章によって構成される．

第 1 章　序論
第 2 章　大規模位置データ分析・流通基盤
第 3 章　大規模位置データの POI 分析手法
第 4 章　大規模位置データ流通のための情報伝播ネットワーク
第 5 章　結論

第 1 章「序論」では，研究の背景と目的，および本論文の構成について述べる．

第 2 章「大規模位置データ分析・流通基盤」では，大規模位置データの管理基盤の中でも位置

データを活用する上で要となる処理である分析手法および流通手法を分類・整理し，それぞれの概要機能について述べる．そして，分析手法，流通手法のおのおのについて，本論文が着目し，研究対象とする範囲を明確にする．さらに，着目する分析手法，流通手法の各範囲における課題について述べ，本論文が解決する課題を明らかにする．

第３章「大規模位置データの POI 分析手法」では，大規模な位置情報が付帯するデータからの POI（Points of Interest）抽出手法について述べる．多くの場合，位置データを扱うアプリケーションにとって，データそのものよりも，POI のほうが重要な場合が多い．POI はアプリケーションによって様々であるが，例えばライフログなどのアプリケーションでは，ユーザが立ち寄ったレストランなどの場所であり，そのような POI は典型的には滞留点（短い時間にデータ点が密集する場所）を検出するアルゴリズムによって自動抽出が図られる．典型的従来手法としては，空間的・時間的変数における閾値を用いたアルゴリズムが用いられ，ユーザが一定時間以上一定範囲以内にとどまったデータセグメントを抽出することで実現される．しかしそのための最適な閾値（パラメータ値）は一般に事前には不明であり，スパイクノイズなどの影響も受けやすい．またデータ点間の距離計算を伴うため大規模データに対してスケールしない．

そこで本論文では，パラメータ設定が容易で，データ数の増大に対してスケールする POI の自動的抽出アルゴリズムについて示す．提案手法は，考慮したい変数（位置情報，時間など）で構成される D 次元ユークリッド空間において定義される LSH（Locality Sensitive Hashing）の持つランダム空間分割特性を応用し，データをランダム空間分割した各小領域に割り当てることで作成したランダムヒストグラムから，あらかじめ設定した所望の特徴指標（滞留点ならばデータ点数）により比較・序列化処理を行うことで，所望の特徴をもった領域を抽出し，POI を特定する．

まず，もっとも基本的な POI として，滞留点やユーザが多くいる場所，つまりデータ点数の密度情報が大きいところとする．そして，その POI の抽出のための提案手法のランダム空間分割特性について理論的に考察し，POI 抽出精度を従来手法と比較することで性能の優位性を示す．また，POI 抽出精度の指定するパラメータの値への依存度がランダマイズ効果によって低減され，トレランス幅が広がることを示す．さらに，ケーススタディを通して実際のログに対して適用したいくつかのアプリケーション例を示し，また実データを用いて計算時間がデータ点数の増加に対して線形依存におさまることをシミュレーションによって示す．

最後に上記の提案手法をさらに一般的なケースに適応可能なように拡張すべく，ランダム空間分割を行うための LSH を一般化する手法と，抽出したい特徴を滞留点からより一般化な指標へと拡張する手法について述べる．LSH の一般化は，データ分布形状が等方的でない場合に，ランダム空間分割に指向性を持たせるものである．そのような指向性を持った LSH の設計理論について示し，指向性をもったデータ分布形状をもつログからの POI 抽出性能を向上することを理論的，実験的に示す．また，抽出したい特徴がデータ点数以外のより複雑な指標であっても，比較指標

を一般化することで同様の手法を適用可能であることを示す．そのアプリケーション例として位置情報が付帯したテキストデータから，機械学習によってテキストを感情分析し，その結果を用いた比較指標を定義することで，感情量の変化というより複雑な POI 抽出条件を定義し，提案するランダム空間分割手法と組み合わせることで，イベント検出への応用が可能であることを確認する．

第 4 章「大規模位置データ流通のための情報伝播ネットワーク」では，位置データのユーザへの効率的な配信流通を分散的伝播技術によって実現する情報伝播ネットワーク構造について述べる．分散的伝搬手法として本論文ではシンプルだが非常に伝搬効率のよいゴシッププロトコルを採用する．実空間上に分布するユーザの間でネットワークを形成し，その上でゴシッププロトコルによって位置データを伝播させる時の伝播効率や伝播パターンを考えるにあたり，情報発信者と受信者の位置関係を考慮することは非常に重要である．位置データは情報発信位置から見て近距離に位置するユーザは高い興味を示す可能性が高いため選択的な情報伝搬が求められるが，同時にそれ以外のユーザの興味は距離に対してニュートラルであることから，ユーザがどの距離にいても効率的に情報が受信できるよう伝搬させることも必要である．両者の要件を実現する情報伝播パターンを実現するには,そのネットワークモデルについて考察することが必要である．

本論文では，効率的な分散ルーティングが可能な Kleinberg のスモールワールドネットワークのトポロジーを，ゴシッププロトコルを用いた情報伝播に応用し，実空間に適用できるようグリッドネットワークから連続空間にモデルを拡張する情報伝搬ネットワークモデルを提案する．そして，そこでの情報伝播特性を解析的手法およびシミュレーションによって検証し，近距離への選択的伝播に加え，発信者からの距離にたいしてニュートラルな効率的伝播を実現することで，提案ネットワークが実空間情報伝播に適していることを示す．まず，提案するネットワークモデルの理論解析により，ロングリンクを通した情報伝播確率が，距離スケールによらない定数を下限にもつことを理論的に示す．また，その伝播特性（システム全体への伝播効率や伝播パターン）を解析的手法およびシミュレーションによって評価する．さらに，実空間上をランダムに動くターゲットを追跡するシミュレーションが可能なシミュレータを構築する．そしてそのシミュレータを用いて，追跡者がネットワークを伝播してくるメッセージを頼りにターゲットを追跡するときの追跡効率を様々なネットワークモデルと比較評価することで，提案するネットワークが情報発信者と受信者の位置関係によらずもっとも効率的に情報取得が可能なネットワークであることを示す．

第 5 章「結論」では，本論文で提案した手法の主たる成果についてまとめ，今後の展開について議論することで，本論文のまとめとする．

以上