

論文の内容の要旨

論文題目 拡散方程式を用いた位置情報 SNS からの

時空間パターンの抽出

– 推薦システムへの応用を目指して –

氏 名 西岡 賢一郎

近年、スマートフォンやウェアラブルデバイスの普及により、自分の行った場所を記録したり、記録した場所を友人などと共有したりできるサービスである位置情報 SNS が使われるようになってきた。位置情報 SNS を使って、施設に行ったことを記録することを、チェックインと呼ぶ。チェックインには、チェックインした施設の緯度・経度などの位置情報が含まれており、このチェックインデータを用いた研究が盛んになってきている。チェックインデータには、時刻と位置に関係したパターンである時空間パターンが含まれており、時空間パターンは、ユーザの将来行動の予測、位置情報 SNS のサービスの質の向上、また位置情報データ自体の管理方法の改善などに有用であるため、チェックインデータからの時空間パターンの抽出は重要である。

しかし、チェックインデータには複数の時空間パターンが重なり合っており、時空間パターンの抽出が難しくなっている。また、チェックインデータは特定時刻に

ユーザが特定位置にいたことを記録しているが、ユーザがチェックインの間にどこにいたか分からないため、スパース性の高いデータとなっており、このスパース性により時空間パターンの抽出がさらに難しくなっている。

そこで本論文では、チェックインデータのスパース性を解決しつつ、定常的なパターン(時刻に影響されないユーザのパターン)、周期的なパターン(周期的にユーザが集まったりいなくなったりするパターン)、突発的なパターン(周期性も定常性もなく、特定日に突然あらわれるパターン)の3つの時空間パターンの抽出方法を構築した。

まず、チェックインデータのスパース性を解決するために、ユーザがチェックインの後に自由粒子として動くことと仮定し、二次元の拡散方程式 $(2\pi t)^{-1} \exp(-|x - y|^2/2t)$ を用いて、ユーザのチェックイン間の位置を時空間分布として推定した。時空間分布は、特定の時刻にユーザがどの位置にいるかを確率分布で表す。このとき、チェックイン時刻に必ずチェックインした位置にユーザがいるように制約を加えた。拡散方程式には、ユーザの移動距離の二乗の時間変化の逆数の次元を持つ物理的な定数であるスケールパラメータが含まれている。ユーザの平均移動速度が遅いとスケールパラメータが大きくなり、ユーザの平均移動速度が速いとスケールパラメータが小さくなるという性質がある。このように、スケールパラメータはユーザの移動速度と関係し、データによって異なってくると考えられるため、データから推定する方法を構築した。スケールパラメータの推定で、最尤推定と逆ガンマ分布を事前分布として用いた推定の二種類を比較し、最尤推定より事前分布を用いたスケールパラメータの推定が外れ値に強く尤度が高い推定であることを示した。

次に、推定した時空間分布を離散化し、PCA (Principal Component Analysis)、ICA (Independent Component Analysis)を適用可能にし、時空間分布から時間成分と空間成分を分解し、全ユーザのチェックインデータから時空間パターンを抽出する方法を構築した。時空間パターンは、時間パターンと空間パターンを分離して抽出し、空間パターンの広がり方や時間パターンの周波数解析によって、それぞれのパターンの特徴について調べた。これにより、全ユーザのチェックインデータから、提案手法を用いて、全体的な傾向を含んだ時空間パターンを抽出できることを示した。

全ユーザのチェックインデータから抽出した時空間パターンでは、全体的な傾向の時空間パターンを抽出することができたが、ユーザが普段生活している範囲によって異なる時空間パターンの抽出ができなかった。そこで、本論文では、1人のユーザに関係した時空間パターンを抽出するために、類似した行動範囲のユーザを集めて時空間パターンを抽出した。

類似しているユーザの定義を、同じ時刻に近い位置にいるユーザとし、各ユーザに対して定義された時空間分布の類似度としてヘリンジャー距離を用いた。ヘリンジャー距離がユークリッド距離であることにより、Ward法が適用可能であるため、Ward法を用いてユーザをクラスタに分け、クラスタごとに時空間パターンを抽出した。各クラスタの特徴的なパターンについては、全ユーザからの抽出した時空間パターンと同様に、空間パターンの広がりや時間パターンの周波数解析をすることで、生活範囲に関係した特徴的なパターンが抽出できていることを示した。

時空間パターンの抽出にて定義した各々のユーザの時空間分布は、ユーザが特定の時刻にどこにいるかを確率分布として表しているため、ユーザの位置の予測にも適用できると考えられる。しかし、定義したモデルのままでは、ユーザの最後のチェックイン後は、時空間分布が広がり続けるため、位置予測として適切ではない。そこで、特定周期の時空間パターンを用いることで、ユーザの最後のチェックイン後の位置の予測をするモデルを構築した。

本論文では、ユーザの時空間分布を使った予測モデル DPM (Diffusion-type Periodic Model) と類似したユーザのチェックインデータを集めて時空間分布を使った予測モデル DPMU (Diffusion-type Periodic Model with similar Users) と時空間分布に対して PCA にて次元削減をした RDPMU (Reduced Diffusion-type Periodic Model with similar Users) の 3 つの予測モデルを構築した。構築したモデルを用いることで、先行研究より複数の評価指標において精度の高い予測ができることを示した。