

ドットデータを基本にした空間データマイニング

愛媛大学工学部 正会員 羽藤英二

1. はじめに

移動体通信システムや RFID タグなどの普及に伴い、実空間上の移動・活動に関する膨大なデータの蓄積が始まっている。これらのデータを施設データなどとあわせて詳細に解析することで、交通計画や管理・運用に活用することが考えられる。

従来の交通需要予測におけるモデル群は、道路ネットワーク上の交通混雑に注目し、アンケートなどで得られる少量のデータを基本にした上で、モデル仮説、検証、モデル決定を経て政策評価に用いられてきた。ここでたとえば都市再生のような問題を考えたとき、都市空間において戦略的に施設の再配置やネットワーク空間の再配分を行うことで、多様な移動・活動パターンを円滑に引き出すことが求められる。この際、スケールの異なる多様で動的なミクロな移動—活動現象に対して、単に適用し既存指標に集約させる従来アプローチは、それだけでは不十分といえよう。

Agrawal(1993)以降、ギガバイトを超える膨大なトランザクションデータを処理する必要がある企業 POS データなどの解析手法として、データマイニングが用いられつつある。本稿では、移動体通信システムや RFID タグなどを利用して得られる膨大かつ正確な位置データ（ドット・データ）の利用を前提に、狭域エリアの移動・活動解析手法として空間データマイニングを考え、その適用可能性について考察したい。

2. ドットデータの基本特性

ドットデータを、都市の 4 次元空間（緯度、経度、高さ、時間）における移動オブジェクト I の位置を示すデータとして以下の式(1)で定義する。

$$Data = \{i, x, y, z, t\} \quad (1)$$

データ解析において、ドットデータを生データのまま使うことはありえない。蓄積されるドットデータは何らかのノイズを含むからである。時空間ネットワーク上のノイズを含む移動軌跡固有の分解能

は、観測機器と分析手法に依存して決定される。

現在交通分野で利用可能なドットデータには、GPS を用いたデータや電界強度の減衰特性、マルチパスパターンを利用したデータ、RFID タグを利用したデータなどがある。

公共交通機関向けの運賃支払い用 IC カードやクレジットカードに用いられている RFID タグ（パッシブタイプ）では、位置精度は 1 m 以内であり、時間や購買履歴がほぼ正確に記録できる。こうしたデータはある特定の意味空間においてイベントが発生したときに観測可能である。

RFID タグは精度の高い位置特定を可能にするが、観測機器の設置が必要となるため、シームレスに活動場所を把握することは難しい。実際に人は時空間ネットワークを動き活動するため、観測機器がないような場所のデータも必要となる。携帯電話やカーナビに搭載される GPS は任意の取得間隔で、こうしたデータを収集することができる。時間精度は原子時計で正確に特定可能であるのに対して緯度経度精度は±10m、高さ方向は±数百 m である。また屋内における精度は緯度経度精度で±数百 m まで低下する。

こうした位置精度のばらつきのため、データマイニング前にドットデータの補正が必要となる。データ補正は人が時空間ネットワーク上で実際に移動可能な領域を限定した上で、位置精度に応じて、その空間領域にいた尤もらしさを尤度として定義し、尤度に基づいて補正を行う。

3. ドットデータのユニット化

補正したドットデータを都市内の狭域エリアの移動・活動パターンの抽出に用いることを考える。分析対象としたい任意の時空間（エリア、経路、特定時間帯など）にドットデータを吸着させ、ドットデータのある行動文脈として解釈可能なデータユニットに集約させる。ここでは図-1 に示すようなデータユニットへの吸着を考える。

キーワード: データマイニング, 回遊行動

連絡先: 〒790-8577 愛媛県松山市文京町3番、089(927)9825、FAX089(927)9843

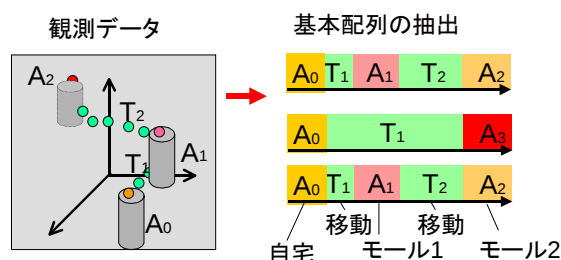


図-1 データ吸着

単位時間あたりの移動距離が(X,Y,Z)で規定される空間位置精度以内であれば、ある空間ユニットに滞在したことを、そうでなければ移動状態にあったことを判定する。この際、図-1 に示す determination cube の中心座標の設定方法は2つある。分析者が任意に設定する場合とドットデータの移動平均を用いて自動的に判定する方法である。この操作によって Data=(ID, X,Y,Z,T)で規定されるデータを(ID,C)として新たに個人IDとdetermination cube毎に格納する。GPS携帯や、プローブビークルを用いた調査において、収集される位置データをそのまま取り扱った場合データ量が膨大になり、そのまま解析することが難しい。determination cube毎に行動文脈を保持したままデータを格納することで、効率的なOLAP(On-Line Analytical Processing)分析が可能となる。

4. 条件抽出の考え方

都市内で頻出する移動・活動相関ルール(Association Rule)を抽出するための計算パラメータの取得概念を図-2に示す。計算配列は個人番号・条件・結論をパラメータとしている。個人番号別に各条件に対する結論の存在する数を該当セルに代入する。変数が正である場合1として個人番号を通して加算し、目的とする条件—結論を満たす該当者のサンプル全員に対する割合を求める。設定した条件・結論に対する確信度・サポートの2つのパラメータが同時に求めることができる。確信度とは、条件部・結論部をともに満たす者が条件部を満たす集合に占める割合、すなわち相関ルールの有効性を表す。サポートは条件部と結論部をともに満たす者が全体に占める割合、すなわちサンプル中にお

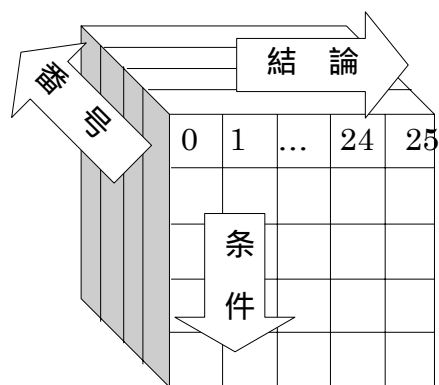


図-2 Association Rule の計算パラメータ取得概念

いてルールが適合できる割合を表す。

5. 計算例

松山の中心市街地における回遊行動データを用いて相関ルールの抽出を行った。図-3は男性と女性の相関ルールの違いを示している。男性と女性で回遊パターンに大きな変化がみられることと、相関ルールに非対称な関係があることが確認できる。こうしたケースでは、従来の分散共分散行列に基づいた分析よりも相関ルールに基づいた分析が必要となるといえる。

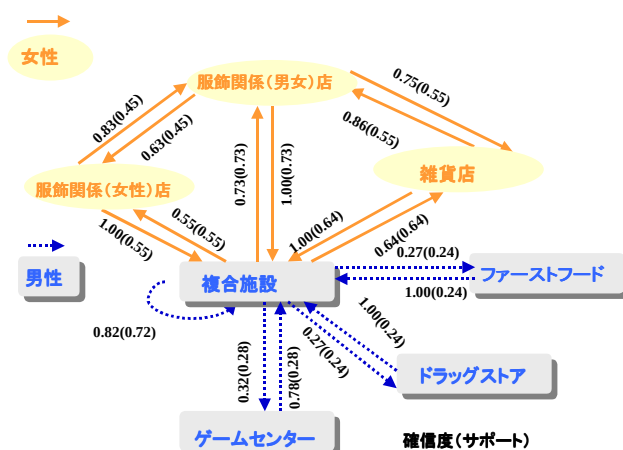


図-3 Association Rule の抽出結果

参考文献

Agrawal, R. et al.(1993) Mining association rules between sets of items in large databases, Proc. SIGMOD, pp.207-216, ACM.