



常を自動で通報したりできると期待される。これはデータ分析の観点では、状況依存の異常検知問題と捉えることができる。

本研究の目的は、交通の状況という潜在情報を、空間移動データを用いてデータ駆動のアプローチで把握することである。交通の研究は交通工学の分野で長く進められてきた。交通工学は流体力学などの知識を応用し、モデルを立てて分析を行ってきた。また渋滞や交通異常の検出には現実の交通を観察して経験的に得られたヒューリスティクスが用いられてきた。本研究はこれとは逆の立場を取り、データの側から交通を理解する課題に取り組む。データを用いて状況すなわちコンテキストを学習することにより、状況依存の異常検知を可能にする。

データに直接現れていない潜在情報の分析は、これまで自然言語処理の分野で活発に研究がなされてきた。文章はデータとしては文字列ないし単語列に過ぎず、そこに含まれる意味内容や話題は通常明示されていない。このような情報を抽出する分析手法が潜在意味解析である。潜在意味解析の手法のなかでも、トピックモデルと呼ばれる潜在変数モデルは拡張の柔軟性が高く、解釈しやすい分析結果が得られることが知られている。本論文は、トピックモデルの考え方を空間移動データに対して適用し、交通状況把握のための潜在変数モデル、および、それを活用した空間移動データ分析手法を提案する。

本論文はつぎの三つの分析事例から構成される。

第一の分析事例では、最も基本的なトピックモデルである潜在的ディリクレ配分法(LDA)を応用した交通状態モデルを提案し、首都高速道路で収集されたプローブカーデータを分析する。「順調」「渋滞」といった交通の状態は直接観測することが困難である。本論文ではこれらの情報を潜在情報、すなわちLDAにおける「トピック」と位置づけ、プローブカーデータからこれらの情報を抽出するための潜在変数モデルを提案する。このモデルによって交通集中渋滞の発生しやすい場所や時間帯がデータから明らかになった。さらに、本論文では交通状態モデルを用いて平常の交通状態と現在の交通状態をそれぞれ評価し、その差異を定量的に評価することで、高速道路上で突発的に発生する交通異常を自動的に検知する手法を提案する。首都高速道路で実際に得られたデータを用いた実験では、個別の車の挙動が異常であるか否かを従来手法と比較してより高速に精度よく識別できることが示された。この異常識別手法をリアルタイムに実行するためのデータ分析基盤の設計および実装についても論じる。

第二の分析事例では、先の分析で用いた交通状態モデルを拡張し、札幌市の一般幹線道路で収集されたプローブカーデータを分析する。札幌市は日本で4番目に人口の多い大都市でありながら豪雪地帯に位置しており、冬季の自動車交通が雪の影響を大いに受ける。そこで本論文では、気象条件を考慮する交通状態モデルを提案する。交通状態はそのときの気象条件に依存すると仮定し、トピックモデルの一つであるパチンコ配分モデルの考え方を応用して潜在変数モデルを設計する。このモデルを札幌市内の実際のプローブカーデータに適用し、気象条件と交通の関係を定性的および定量的に評価する。

第三の事例では、スマートフォンを用いて収集された人の移動データの分析を試みた。プローブカーデータと異なり、スマートフォンのデータは交通移動手段が明らかでない。本論文では交通移動手段を潜在情報と位置づけ、データとして観測される移動方向や所要時間と交通移動手段の関係を記述する潜在変数モデルを提案する。中国・北京市において実際に得られたスマートフォンの位置情報データを用いて行った実験では、このモデルによって交通移動手段と人の動きの関係が学習できることが示された。このモデルを用いて、出発地・到着地・所要時間のデータからその間の経路を補間する実験も行い、4分の3以上のデータが正しく補間できる結果を得た。潜在情報を活用することでデータの補間や予測が効率的に実行できる可能性が示された。

本論文は、空間移動データに対する潜在変数モデル応用の方法論を示し、実際の分析事例を通じて、このアプローチでの交通状況把握の有効性を実証した。本論文がとったデータ駆動のアプローチは知識駆動のアプローチと対立するものではなく、むしろ共存するものである。潜在変数モデルは柔軟性が高いため、より高度なデータ分析に向けて、交通に関して得られている知見をモデルに追加し、現実の分析課題に対してより効果的と知られているデータや特徴量を用いることも可能である。技術の進展と社会の高度化に伴い、データの種類や分析課題は今後さらに多様化・複雑化していくと予想される。そのような中においても、潜在変数モデルは、複雑な人間社会をデータから理解するための有効な手段として活用されていくと予想する。