

適用分野を広げる データマイニング技術

ユビキタスデータからの 知識発見に向けて

多くの人や物にセンサが搭載され、そこから大量のデータが収集されるユビキタス時代では、これらのデータの中に隠された知識や傾向を発見して意志決定などに役だてることが重要な応用になると思われます。

大量のデータから隠れた知識を発見する技術としては、データマイニング技術がよく知られていますが、東芝では、ユビキタスデータの特徴である時空間的要素を持つデータのマイニング技術として、「空間データマイニング技術」や、リアルタイムに発生する時系列データを分析する「リアルタイムデータマイニング技術」を開発しています。

従来のデータマイニング技術

データマイニング技術とは、大量のデータから隠れた知識を発掘する技術で、流通業におけるPOS (Point Of Sales) データの分析などでよく使われています。例えば、米国のスーパーマーケットで「紙おむつと缶ビールが同時に購買される」などの傾向を発見し、両製品の陳列位置を近づけることで売上げを増大させた成功事例などが報告されています。

来るべきユビキタス社会では、人や物がネットワークでつながり、あらゆる場所でリアルタイムに発生するデータが増大するので、これらのデータを有効利用するためのデータマイニング

方式が必要となります。

ここでは、このような位置情報やリアルタイムの時系列情報を持つデータのマイニング技術を紹介いたします。

空間データマイニング技術

位置情報を持つデータのマイニング技術に関しては、地図などの実世界空間上に広がるデータから隠れた知識を発見する「空間データマイニング技術」の開発を進めています。

空間上に広がるデータを分析すると、「どのような場所で何が発生するか」といった場所に依存した知識が得られ、位置を特定した効果的な対策ができるようになります。

この技術の全体像を図1に示しま

す。この技術では、「分析対象データに対して、地図上のどんな物体が、何個くらい、どのくらいの距離で近接する傾向にあるか」といった従来は発見困難であった詳細な近接関係や、「どの辺りに、どんな条件の分析対象データが密集する傾向にあるか」といったデータの密集状態とその条件を発見できるようにしました。

この技術を組み込んだ空間データマイニングシステムの画面例を図2に示します。この空間データマイニングシステムは、分析対象のデータを地図上に表示する部分とデータマイニング結果を表示する部分で構成されます。

このシステムを用いると、立地条件を考慮した店舗の売上分析や、土木建

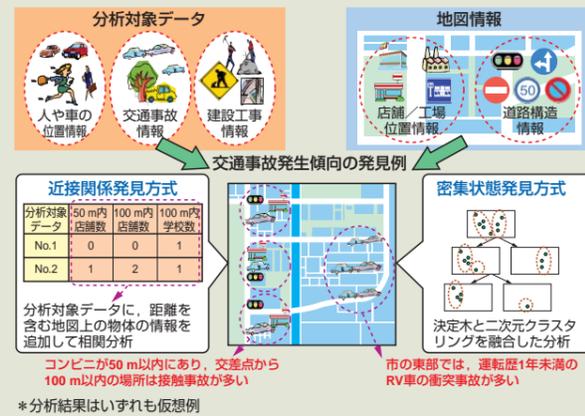


図1. 空間データマイニングの概要 — 分析対象とするデータと地図上の情報との近接傾向や、データの密集傾向などを発見します。

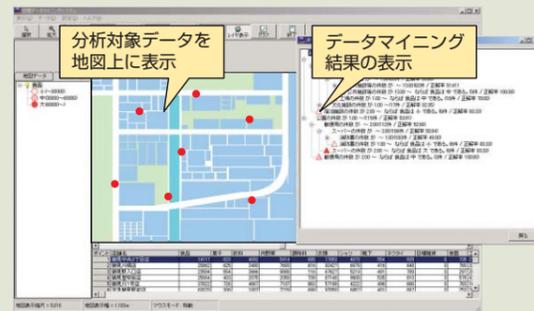


図2. 空間データマイニングシステムの画面例 — メインウィンドウ上に地図が表示され、分析対象のデータがプロットされます。データマイニング結果は別ウィンドウに表示され、ウィンドウ内で詳細を確認することができます。

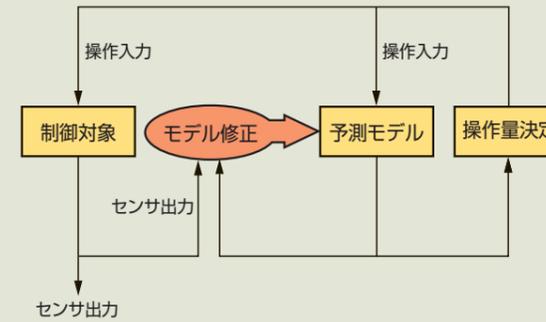


図3. リアルタイムデータマイニングの概要 — 目標とするセンサ値を予測しながら制御し、予測が外れ始めるとデータマイニングを再実行して予測モデルを修正し、精度の良い予測と安定した制御を実現します。

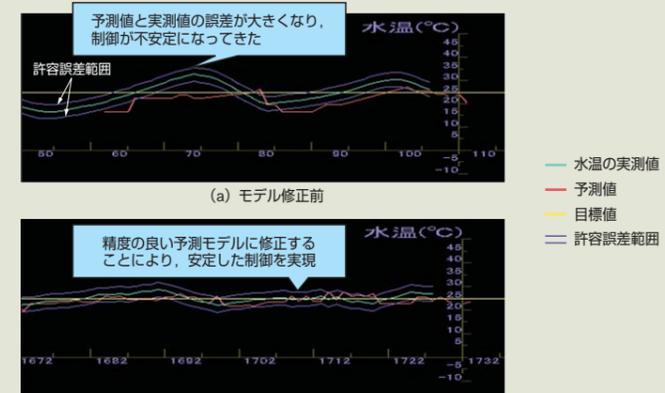


図4. リアルタイムデータマイニングシステムの画面例 (水温制御へ適用した場合) — モデル修正前と修正後と比較すると、予測精度や制御の安定性が向上していることがわかります。

設などの公共工事の最適計画支援などを行うことができます。

リアルタイム データマイニング技術

時系列データの分析に関しては、リアルタイムに発生するデータを分析し、時々刻々と変化する傾向を抽出する技術を開発しています。

工場などの製造現場において、未来のセンサ値を予測して、そのセンサ値に対する最適な制御を行うことは重要な応用です。一般に、対象システムが複雑になると数式による厳密なモデル構築が困難となるため、計測されるデータから近似モデルを構築する方法を採ります。この際に、リアルタイム

データマイニング技術を用いて、予測モデルから得られる予測値と製造装置から得られる実測値をオンラインで比較し、予測誤差が大きくなった場合に、自動的にモデルを再構築し、精度の良い予測を実現するシステムを開発しました。

図3は、リアルタイムデータマイニングを予測制御に適用する場合の構成です。予測モデルの出力を参照しながら操作量を決定して対象を制御すると同時に、制御対象のセンサ出力と予測モデルの出力が合わなくなると、モデル修正機構により予測モデルを再構築します。

図4は、水温制御に適用した例を表しています。緑線は水温の実測値、

赤線は予測値、黄線は目標値、紫線は許容誤差範囲を示しています。初めのうちは図4(a)のように予測値と実測値が合わず、その結果、水温は目標値から外れていますが、予測モデルの自動修正により図4(b)のように水温が目標値近くに制御されるようになります。

この技術は、大規模プラントの制御をはじめとして、生体や自動車などに搭載されたセンサデータから行動パターンを発見する応用などにも有用であると考えられます。

今後の展望

現在開発中の空間データマイニング技術、リアルタイムデータマイニング技術を融合し発展させ、「いつ、どこで、何が起るか。その原因は何か」といった時空間的な知識を発見する、「時空間データマイニング方式」を開発予定です。この技術を用いると、人間の行動における移動時間や移動場所と体調との相関関係を用いた健康管理や、自動車の運行履歴を用いた省エネ運転アドバイスなど、新しい応用が可能となります。

仲瀬 明彦

研究開発センター
システム技術ラボラトリー研究主務