

平成14年度 情報工学コース卒業研究報告要旨

坂部 研究室	氏 名	伊 藤 陽 生
卒業研究題目	データマイニングのための柔軟なデータ取得・統計を支援するAPIの設計	
<p>近年の計算機の高速化、大容量化に伴い、これまでは収集されることのなかった大量のデータを解析し、様々な研究やマーケティングに利用するデータマイニングが可能となる環境が整ってきた。しかしながら、SQLによる問合せを行う既存のデータベースでは実行可能な操作が限定されており、思い通りのデータを抽出するにはかなりの工夫と労力を要する。</p> <p>本研究では、これらの問題を解決するために、データマイニングシステム構築の基盤となるAPIの設計を行う。また、これをJavaで実装し、評価した。設計したAPIは、</p> <ul style="list-style-type: none">• 行と列から構成される表を扱うデータのクラス• 表を操作するためのインタフェース群• 表を操作するためのクラス群(上記インタフェース群の実装として)• 表と様々な情報源との間でデータをやりとりするための入出力クラス群 <p>の概念から成り、データマイニングのためのプログラムを容易に作成することができる。表を操作する手段として、予め用意されたクラス群を利用する他に、ユーザによってインタフェースの実装を行なうことで必要なクラスを作成し、利用するという方法がある。これにより基本的な操作を容易に行ないつつ、柔軟なデータ操作が可能になる。またJavaの豊富なクラスライブラリと連携することで、アプリケーションやWebサービス等への応用が容易になるという利点もある。</p> <p>応用事例として、小売店のレシートデータを解析するプログラム、並びにウェブサーバのアクセスログを解析するプログラムを作成する実験を行なった。その結果、効率よくプログラムが作成でき、これらの作成のためにこのAPIが十分な機能を持っていることが分かった。また処理の流れを手続き的に記述できるため、同様の処理をSQLで行なった場合に比べソースの可読性を高く保つことができた点が優れている。テーブルの作成時以外は情報源のフォーマットを意識せずに開発が行なえた点が開発効率の向上に寄与していると考えられる。しかしながら、現段階では、データ操作メソッドの実行速度が極端に遅い、という問題点も明らかになった。</p>		