

金融業におけるデータマイニングの理論と実務 ～不動産賃料の推定モデル構築～

小野 潔
三菱東京 UFJ 銀行
中小企業部

Kiyoshi Ono
Small & Medium Enterprise Banking Division,
The Bank of Tokyo-Mitsubishi UFJ,Ltd.

キーワード： データマイニング，不動産賃料モデル，決定木，EM アルゴリズム

要旨

本報告は、金融業のデータマイニングの理論を説明し、次にデータマイニングによる2種類の不動産賃料モデルの構築法を解説する。

データマイニングは機械学習を用いてデータウェアハウスから意味ある情報（知識、パターン、ルール）を抽出する技術である。データベースから知識発見は、KDD プロセスから成り立つ。データマイニングの分類器には、決定木、アソシエーションルール、ニューラルネットワーク、SVM、MBR、GA/GP 等がある。学習にはコミュニティ学習、アンサンブル学習があり、そのほか、分類器を組み合わせたハイブリッドモデル、欠損値を捕捉するEMアルゴリズムが、データマイニングのモデル構築に利用される。

データマイニングによる金融業の最初の成功事例は94年米国の与信審査モデルである。日本は97年から本格的な導入が始まった。現在ではリテール金融業の発展の原動力になっている。しかし業務でモデルを成功させるのは容易でなく、分析結果を業務に適用させるビジネスモデルが必要となる。

日本不動産金融工学学会主催の賃料分析コンペティションにおいて、“モデルの可読性”と“モデルの精度”の観点から2種類のモデルを構築した。精度の観点”では推定モデルの誤差を最小限にすることに重みをおくが、“可読性の観点”では利用者のモデルへの理解度が重要になる。①可読性のすぐれたモデルは9個の構造ツリーと割高/割安表から構成され、やや精度が劣るが業務の専門家にわかりやすいモデルを構築できた。②EMアルゴリズムを応用して「駅ごとの部屋タイプ別の賃料と面積」データの欠損値を捕捉することで、モデルの精度の向上に成功した。