

平成 27 年度 TSUBAME 産業利用トライアルユース 成果報告書

利用課題名 中古マンションの将来価格予測  
英文: Future Value prediction of condominium

利用課題責任者 伊藤 嘉道  
Yoshimichi Ito

所属 株式会社東京カンテイ  
Tokyo Kantei Co., Ltd.  
URL <http://www.kantei.ne.jp/>

我が国の住宅ローンでは、住宅価格が下落しオーバーローンになってしまうと住宅を買い替えることが難しい。そこで、当社では、市場流動性の高い大都市圏の中古マンションについて、「残価設定型住宅ローン」を提案している。「残価設定型住宅ローン」を実現するためには、中古マンションの将来の価格の予測モデルの構築と検証が必要である。精度の高いモデルを構築するには、膨大な回数のシミュレーションが必要となる。このため、本課題では、TSUBAME を利用して、モデル構築を行い、最適な並列演算の方法と並列数を探った。

At the Japanese housing loan, a borrower faces difficulty to purchase a new house when the price of the house one's currently own becomes lower than the housing loan balance by the decline in the housing prices.

To give one of countermeasures for this subject, we are proposing a new scheme "residual value set mortgage". This is especially for the condominium in the metropolitan area because of containing a large number of deal.

We have come to an idea that we need to construct and verify the predictive model of future prices of the condominium to realize this scheme. It requires to simulate vast number of times to secure high accuracy.

Therefore, we explored the optimized method of parallel arithmetic and number to construct a predictive model by using TSUBAME.

*Keywords:* 不動産, 将来価格予測, 局所線形回帰分析, 変数選択

### 背景と目的

我が国の住宅ローンでは、住宅価格が下落してローンの残額を下回ると、住宅を買い替えることが難しい。そこで、当社では、市場流動性の高い大都市圏の中古マンションについて、貸し手と借り手の双方に住宅価格の下落リスクを分散させる、「残価設定型住宅ローン」を提案している。

「残価設定型住宅ローン」は、融資を、リコースローン部分とノンリコース部分に分ける。リコースローン部分は、返済期間に償却する部分であり、毎月元利を支払う。ノンリコース部分は、返済期間満了時の残価の部分であり、毎月金利のみを支払う。

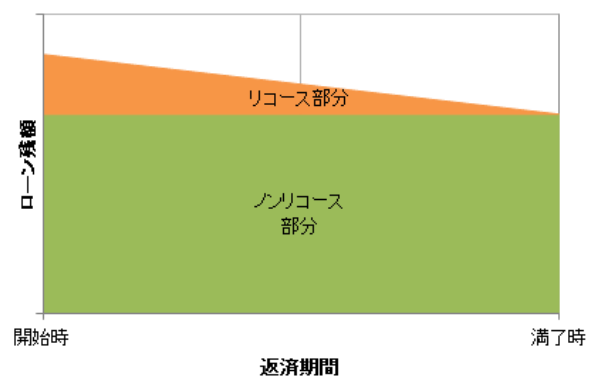


図 1 残価設定型住宅ローンの返済イメージ

これにより、借り手は、自動車の残価設定クレジットのように、毎月の支払い額が少なくても、より条件の良い住宅に住むことができ、返済期間満了時にはノンリコース部分の返済義務を負わずに住宅を手放すか、再

度ローンを組むか選択することができる。

一方、貸し手は、返済期間満了時に住宅価格がノンリコース部分の元本を下回った場合は、その不足分についてリスクを負わねばならないため、将来底割れしないであろう価格を定量的に予測できなければならない。

将来の中古マンション価格は、価格時点を将来に置いた DCF 法により求めることを提案している。収益還元法の一つである DCF 法で価格を求めるのには、対象マンションを賃貸した場合の賃料と、過去時点の予測賃料を基に過去の取引事例を基に求められた割引率により計算できる。そのため、マンションの賃料の予測モデルが基礎となる。

そこで、マンション毎の賃料の予測モデルを構築し、経年と賃料の関係を求める。経年と賃料の関係は特定の関数の関係とは限らないため、ノンパラメトリックな手法の一つである局所線形回帰を用いる。また、同じマンションでも部屋毎、時代毎に賃料形成要因が異なると考えられるために、局所毎に情報量基準(BIC)が最小になるような変数の組み合わせを選択する探索的な変数選択(変数増減法)を行う。

そのため、経年を 1 年毎に 60 個、専有面積を 20~120 m<sup>2</sup>を 2 m<sup>2</sup>刻みで 50 個、事例発生年を 1 年ごとに 20 個の局所を設定すれば、60×50×20=60,000 個の局所モデルが必要となり、10,000 棟のマンションであれば、60,000 局所×10,000 棟=6 億個のモデルが必要となる。さらに、局所毎に変数選択を行うと、説明変数の候補が 10 個の場合、それぞれの局所で 1,024(2<sup>10</sup>)回の重回帰計算を行うことになる。将来の中古マンション価格の信頼性は、マンションの賃料の予測モデルの精度に依存し、精度の高いモデルを構築するには、膨大な回数のシミュレーションが必要となる。従って、本課題では、中古マンションの将来の価格の予測モデルの構築と検証を目的として、最適な並列演算の方法と並列数を探ることとする。

**概要**

弊社既存のマンションの賃料予測モデル構築システムを元に、並列演算等の機能追加を行い、実行環境の条件を変えながら、並列演算のパフォーマンス測定を行った。研究開発段階であることから interpreter で各

種パッケージの豊富なR言語を使用した。

モデル計算は、マンションごとに局所線形回帰で構築し、経年・専有面積の 2 項目で局所を構成する。さらに、変数選択を各局所で行う。

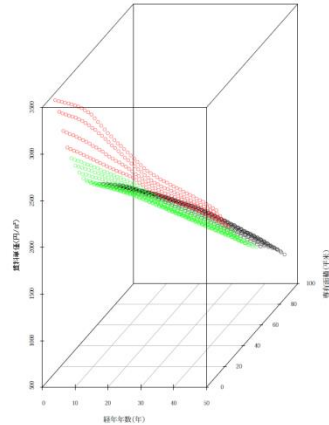


図 2 専有面積・経年ごとの賃料単価のイメージ

試行では、並列演算の計算粒度と並列数、また、GPU 利用の有無の組み合わせで実行時間を測定した。なお、計算粒度を棟としたものは、計算時間の推測が容易なため、試行しなかった。

表 1 試行内容

サンプル数	70 万件
説明変数の数	12 個

表 2 並列処理の計算粒度

粒度	特徴
変数選択	<ul style="list-style-type: none"> <li>汎用性が高い。</li> <li>オーバーヘッドが大きい。</li> <li>ノルムによるウェイト計算が並列化できない。</li> </ul>
局所	<ul style="list-style-type: none"> <li>ノルムによるウェイト計算を並列化できる。</li> </ul>
棟	<ul style="list-style-type: none"> <li>負荷の不均衡が起こる。</li> <li>オーバーヘッドが小さい。</li> </ul>

表 3 試行内容

計算粒度	<ul style="list-style-type: none"> <li>変数選択</li> <li>局所</li> </ul>
並列数	<ul style="list-style-type: none"> <li>1node * 1,2,4,6,8,12cores,</li> <li>2node * 12core</li> <li>4node * 12core</li> </ul>
GPU	<ul style="list-style-type: none"> <li>利用しない</li> <li>R gputools パッケージ[1]を利用</li> <li>R gmatrix パッケージ[2]を利用</li> </ul>

## 結果および考察

### (1) 変数選択の並列演算化

ステップワイズ法による変数選択部分を並列化して試行したところ、ノード間の通信のオーバーヘッドが大きく、極めて遅くなった。これは、並列演算の内容が重回帰分析のみであるのに対し、サンプルが大きく転送時間が長いからである。

なお、R の並列計算の特徴として、メモリ共有を扱えないため、同一ノードであっても、データ転送が発生する。

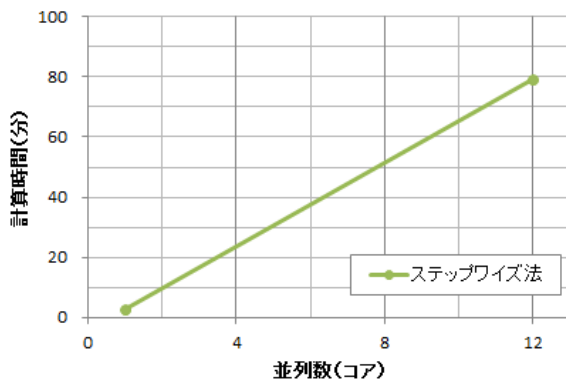


図 3 並列数と計算時間

### (2) 局所の並列演算化

局所計算の並列化はパフォーマンス改善に良好な結果であり、複数ノードへ計算を分散させてもなお高速化の効果が見られた。これは、1 個の局所に対して、データ転送は 1 回であるのに対して、並列演算の内容はサンプルのノルムによるウェイトの計算と変数選択による重回帰分析を含み、計算時間が十分長いからである。

なお、参考までに、強制投入法により変数選択を行わなかった場合は、48 コアによる並列演算では、計算時間が長くなった。これは、本プログラムでは重回帰分析より、ノルムによるウェイトの計算の方が計算時間が長いという特徴が影響している。

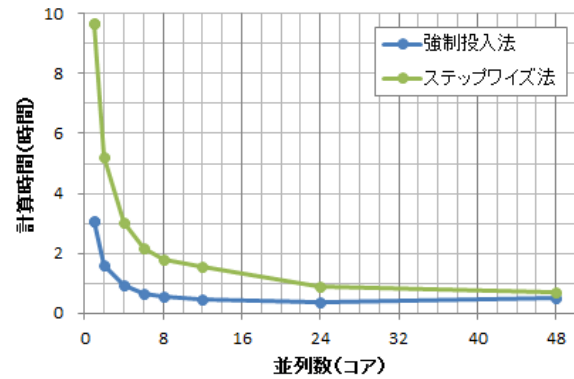


図 4 並列数と計算時間

### (3) GPU によるベクトル計算の高速化

GPU による高速化の検証は、効果の規模感が不明であったため、プログラムから重回帰分析の部分のみを切り出し、GPU の利用方法による効果の差異を検証した。

その結果、gputools パッケージでは計算時間が長くなった。理由は、gputools は、演算・関数の一つ一つで CPU メモリと GPU メモリの間で往復のデータ転送を行うためであり、例えば以下の式では、solve,crossprod2 回,%\*\*%の計 4 回データ転送が行われ、単純な計算あるいはその組み合わせは、オーバーヘッドが大きいからである。

```
solve(crossprod(ev)) %**% crossprod(ev, rv)
ev: 説明変数、rv: 目的変数
```

一方、gmatrix パッケージでは、計算時間が約 50% と短くなった。これは、GPU メモリ上のデータを R の変数として扱えるため、計算結果を GPU メモリに置いたままそれを用いて次の計算を行えるため、転送によるオーバーヘッドが小さくなることによる。また、QR 分解による説明変数のランク落ちの確認も計算時間が約 15% と良好な成績であった。

gmatrix は、GPU メモリ上の変数を続けて使用できるため、重回帰だけでなく、その前後のロジックも GPU で連続して扱うことで、プログラムの広い範囲の高速化が期待できる。

っていきたい。

#### 参考文献

- [1] <https://cran.r-project.org/web/packages/gputools/> Josh Buckner らによる。
- [2] <http://epbiwww.case.edu/index.php/people/faculty/97-morris/> Case Western Reserve University の Nathan Morris 助教授による。

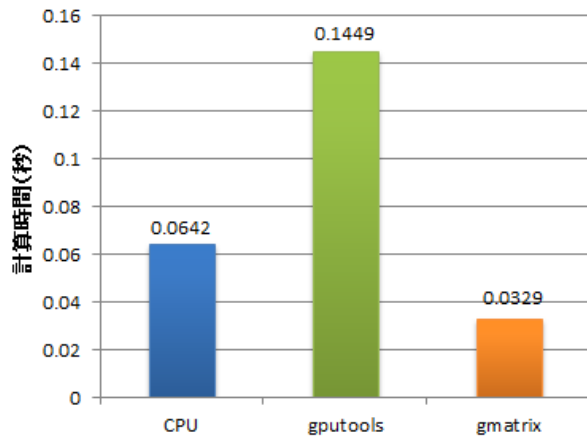


図 5 重回帰分析の計算時間

※サンプル 70 万件、説明変数 12 個の計算結果の平均

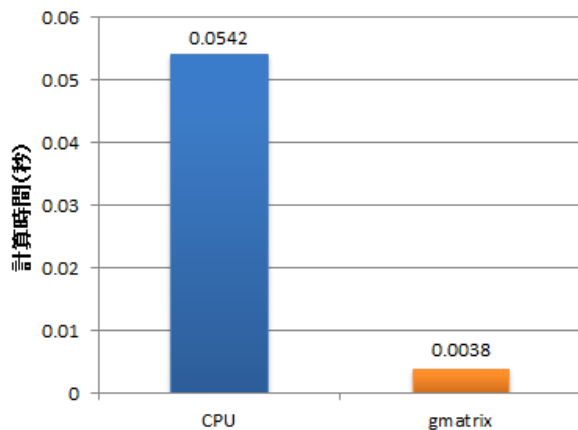


図 6 QR 分解の計算時間

#### まとめ、今後の課題

今回は、TSUBAME により、中古マンションの将来の価格の予測モデルの構築において局所の粒度における並列演算の有効性を確認できた。

また、GPU による並列演算の適用の可能性あることを確認できた。

今後は、モデル構築後の DCF 法による将来の中古マンション価格の推測における並列演算の適用や、GPU のより広い範囲での適用、GPU の並列演算と CPU での並列化の組み合わせについて、可能性を探