

図書館における最適蔵書管理の研究

2018SS063 高橋颯汰 2018SS065 田村仁

指導教員：鈴木敦夫

1 はじめに

本研究は、南山大学図書館に所蔵されている本の蔵書管理問題をオペレーションズ・リサーチの手法を用いて解決することを考える。

現在、南山大学の図書館には全体で 70 万冊以上の本が所蔵されている。図書館の収納スペースは常に限界に近い状態であり、図書館に入りきれない本は外部の倉庫に移されている。しかし、所蔵数は毎年増加しており、このままでは新しい倉庫を借りることや、図書館の増築をすることが必要になる。さらに倉庫へ移動させる本を選定するのにも明確な判断基準がなく、仮に倉庫へ移動した本を借りたりしたいという人が来た場合、図書館へ運搬する費用と時間を要してしまう。本研究ではこの状況を改善するために、本を図書館に維持、保管庫に移動、処分の 3 つに割り当て、オペレーションズ・リサーチの手法を用いて倉庫の維持費や処分の費用の最小化を目的としている。

本を割り当てるシステムを作成するにあたり、図書館の本が将来的にどれだけ借りられるかを数値として求める必要がある。そこで、その数値を求める手法として統計解析ソフト R による回帰分析と、プログラム言語 Python で利用できる Prophet による時系列予測を使用した。回帰分析では、類と呼ばれる日本十進分類法で定められた本の種類ごとに分析を行い、本がどのような傾向で借りられなくなるかを求めた。また、Prophet と呼ばれる機械学習では各本一冊ずつの貸出履歴から将来の需要を求めた。

問題解決にあたっては、オペレーションズ・リサーチの手法である 0-1 整数計画法で定式化を行った。その式に上記 2 つの手法を用いて得られた数値を代入することで費用の最小化を目指す。また、最適化計算には Python の MIP を使用しており、CSV 形式のファイルに結果を出力する。また、図書館における維持費や処分にかかる費用については、費用の比率によって結果がどのように変わるかを求めた。

2 使用データ

本研究で利用するデータとして、大学図書館から過去 20 年間の貸出履歴を提供してもらった。このデータにはその本が図書館に所蔵された検収日から、各本が借りられた日付と時間、請求番号、配置場所が「0,1 類」「2,4,9 類」「3 類」「5,8 類」「6 類」と分類ごとに分けられて記録されている。

今回の研究では、検収日が 2001 年以降である本の検収日と借りられた日付を抽出して扱う。2000 年以前に所蔵された本を使用しない理由に関しては、貸出履歴のデータが

少ないことと、その本が所蔵された時期からどのようにして借りられたのかという全体のデータが無いためである。

3 問題解決へのアプローチ

図書館に所蔵されている本は検収日から年を追うごとに借りられなくなっていくことが古くから知られている [1]。そこで、南山大学図書館の本も同様の傾向で借りられなくなるのかを、統計分析ソフトである R での回帰分析と、プログラム言語の Python で利用できる Prophet で時系列予測を行って調べた。

3.1 回帰分析

図書館から提供してもらったデータを元に、各書籍が年ごとに何冊借りられたのかを Excel でまとめ、各年と翌年の貸出回数を用いて統計分析ソフト R で単回帰分析を行った。しかし、どの本も貸出回数が少なく、決定係数の値が小さかったため、傾向が読み取れなかった。よって、このデータを元に以下のような処理をした。

- (1) 1 年ごとに貸出回数を分けていたものを、4 年おきの年代にまとめる。
- (2) 各年度の前年度と次年度の値を用い、移動平均を求める。

これらの処理を行ったデータが以下の図になる。

	A	B	C	D	E
1	ID	検収年	ave2	ave3	ave4
2	852873	2001	7.33	6	1.33
3	852874	2001	3.33	1	1.33
4	852876	2001	1.33	1	1.33
5	852877	2001	3.67	2.33	1.33
6	852878	2001	1.67	1.33	0.33
7	852883	2001	5	6.33	4.67
8	852885	2001	5	7.33	6.67
9	852888	2001	3.33	2	0.33

図 1 R に読み込ませたデータ

これらの処理を行ったことにより、回帰モデルを導いたところ決定係数の値は各類ごとに

- 「0,1」……0.8379
- 「2,4,9」……0.8076
- 「3」……0.8563
- 「5,8」……0.6669
- 「6」……0.8639

となり、「5,8」類以外は高い数値を得ることができる。図 2 は第 3 類の回帰分析をプロットした図である。

- また、この時の各類の 1 次項の係数は、
- 「0,1」……0.73795

「2,4,9」 ……0.697264
 「3」 ……0.711086
 「5,8」 ……0.396744
 「6」 ……0.7430

となり、こちらも「5,8」類以外の大学図書館の本の貸出回数は、4年おきに約3割ずつ減少していくことが分かった。

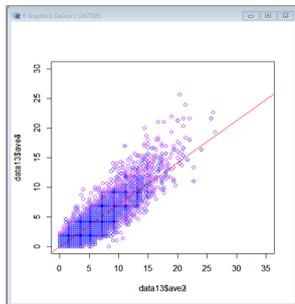


図2 第3類の移動平均を考慮した単回帰分析

3.2 機械学習による予測

Prophet とは Facebook が公開している OSS ライブラリで、Python によって比較的容易に時系列予測が行えるものである。

大学図書館から提供してもらったデータを元に、下図のように毎月の貸出回数をその本の検収月からまとめる。下図の場合、2003年5月には0回、6月には2回借りられていることになる。Prophet の仕組みとして、日付まで入力する必要があるので 2003/5/1 となっているが、実際は月の貸出回数がまとめられている。

2003/5/1	0
2003/6/1	2
2003/7/1	1

図3 Prophet に読み込ませるデータ

一例として、実際に Prophet による需要予測の結果を提示する。図4は第3類において過去20年間で最も貸出回数が多かった分類番号：0868947の書籍についてのプロット図である。

このように貸出履歴から、その本が将来的にどれだけ必要があるかをプロット図と数値で示してくれる。しかし、本来 Prophet は連続性のあるデータに用いるため、図書館の貸出予測のように離散的なデータを使った予測には用いにくいと考えられる。

そのため、2019年までの貸出履歴から2020年の貸出予測を行い、予想と実際の結果を照らし合わせることで、本研究で Prophet による機械学習を用いても良いかを確かめた。

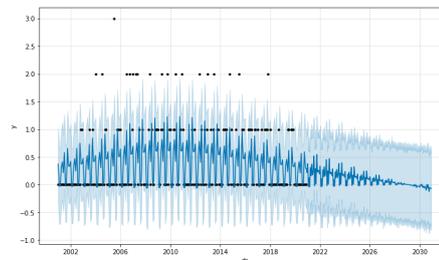


図4 分類番号：0868947の貸出実績と予測のプロット図

その結果、貸出回数が0回の本100冊で予測した場合の絶対誤差は0.008171095。貸出回数が1回以上あった本100冊での予測では1.6926514という値をとったため、誤差が小さいことから本研究で使用することにした。

3.3 数理モデルへの応用

今回の目的関数で使用するものは、回帰分析で求めた1次の項の係数と、Prophet で得られた予測値の上限である yhat upper の値である。それぞれ別の目的関数を用い、本の分類を行う。

また、単回帰分析は各分類ごとに1次の項の係数を求めているので、類の本全てを一度に分類できるが、Prophet は仕組み上、一冊ずつしか分析することができない。そこで、Prophet を用いて一度に複数冊の分析ができるシステムを作成したが、それでも全ての本を分析するのに膨大な時間を要するため、今回の研究では一部の本を用いて分類を行う。

4 0-1 整数計画法による本の割り当て

4.1 0-1 整数計画問題

「図書館に維持」、「倉庫に移動」、「処分」の3つの分類における本の割り当て問題として、倉庫の維持費や処分にかかる費用の合計が最小になるような0-1 整数計画問題を解く。0-1 整数計画問題は、すべての変数の値が0と1しかとらない整数制約がある問題である [2]。この問題では2つの変数を用いてそれぞれの分類について割り当てるか割り当てないかを決定しているため上記の解法を用いる。

4.2 分析結果を用いた問題の解法

次の期間の本の割り当てのモデルは次のようにする。

目的関数を図書館および保管庫に本を置く維持費、本の運搬費用、処分費用の合計の最小化とする。

回帰分析を用いた目的関数では、本を処分または保管庫に移動させてしまった後にその本が借りられてしまう場合のリスク費用は最後に借りられた期間の貸出回数と回帰直線の傾きにより求める。また、回帰直線の傾きに本が何期間前に借りられたかを乗じることで、本が最後に借りられた時点から予測を行う。

また、Prophet を用いた目的関数では、貸出回数の予測値の95%の上限値を各リスク費用に乗じることで分類を

行う。

制約条件としては4つの条件が考えられる。1つ目は図書館の利用者の多くの需要に応えるために、図書館に置かれる本は少なすぎたはいけないので、図書館に置かれる本の大きさの合計が図書館の容量の8割以上かつ上限を超えないようにすること。2つ目は本を多く処分しすぎて現在借りている保管庫に本がほとんど入っていない状態になることを防ぐために、保管庫に置かれる本の大きさの合計が保管庫の容量の4割以上かつ上限を超えないようにすること。3つ目は本を処分するか、図書館または保管庫に置くかのいずれかに割り当てるようにすること。4つ目は変数をバイナリ変数とすることである。

4.3 定式化

記号の定義

I : 本 i 全体の集合

c_1 : 図書館に本を置く維持費

c_2 : 保管庫に本を置く維持費と運搬費用

c_3 : 処分費用

a_i : 本 i の大きさ

g : 図書館の容量

h : 保管庫の容量

R_i : 本 i を保管庫に移動したときのリスク費用

r_i : 本 i を処分したときのリスク費用

m : 回帰直線の傾き

n_i : 本 i が最後に借りられた期間の貸出回数

d_i : 本 i が最後に借りられたのは何期間前か

Y_i : 本 i の \hat{y} の値

変数

$$x_i = \begin{cases} 1 & \text{本 } i \text{ が図書館に維持される} \\ 0 & \text{本 } i \text{ を図書館に置かない} \end{cases}$$
$$y_i = \begin{cases} 1 & \text{本 } i \text{ を保管庫に移動させる} \\ 0 & \text{本 } i \text{ を保管庫に置かない} \end{cases}$$

目的関数

$$\min. \sum_{i \in I} \{c_1 x_i + y_i (c_2 + n_i m^{d_i} R_i) + (1 - x_i - y_i) (c_3 + n_i m^{d_i} r_i)\} \quad (1)$$

制約条件

$$0.8g \leq \sum_{i \in I} a_i x_i \leq g \quad (2)$$

$$0.4h \leq \sum_{i \in I} a_i y_i \leq h \quad (3)$$

$$1 - x_i - y_i \geq 0 (i \in I) \quad (4)$$

$$x_i \in \{0, 1\} (i \in I) \quad (5)$$

$$y_i \in \{0, 1\} (i \in I) \quad (6)$$

数式の説明

- (1) 単回帰分析の値を用いた目的関数として、本を図書館および保管庫に置く維持費、本の運搬費用、処分費用の合計を最小にする..
- (2) 図書館に置かれる本が少なすぎたはいけないので、図書館に置かれる本の大きさの合計が図書館の容量の8割以上かつ上限以下におさめる。
- (3) 本を多く処分しすぎたを防ぐために、保管庫に置かれる本の大きさの合計が保管庫の容量の4割以上かつ上限以下におさめる。
- (4) 本 i は図書館または保管庫に置かれるか、処分するかのいずれかに割り当てる。
- (5) (6) 変数 x_i, y_i はバイナリ変数。

Prophet の予測結果を用いる場合は、目的関数を以下のようにする。

$$\min. \sum_{i \in I} \{c_1 x_i + y_i (c_2 + Y_i R_i) + (1 - x_i - y_i) (c_3 + Y_i r_i)\} \quad (7)$$

なお、制約式は同様である。

4.4 本の自動割り当てシステムの概要

問題を解く際には、20年間の貸出履歴を4年ごとに分けた合計5期間分の貸出回数と、その本が最後に借りられたのが何期間前か示す数値が必要になる。そのため、図書館に提供してもらった貸出履歴から各本が5期間ごとに何冊借りられたかをCSVファイルに出力するプログラムを作成した。また、そのCSVファイルを、Excelの機能を用いて最後の貸出回数と、それが何期間前かをまとめた2つのCSVファイルに分けた。本の割り当てを行うプログラムではこの2つのCSVファイルを使用する。そのCSVファイルをもとに目的関数を最小とする0-1整数計画問題をPythonのMIPを用いて作成したプログラム上で計算し、「図書館に維持」、「倉庫に移動」、「処分」の3つの分類に本を割り当て、実行結果をCSVファイルに3つ分類それぞれの合計冊数とともに出力するというシステムになっている。MIPとは、Python上で上記のような数理最適化問題の目的関数や制約式をプログラムに記述することで、容易に最適解を求めることができるツールである。

5 割り当て結果

回帰分析の結果に基づいた予測

実際に第3類の47820冊の貸出回数のデータを用いてプログラムを実行した。今回使用する回帰直線の傾きは、第3類の4年毎の貸出回数の移動平均のデータを単回帰分析して得られた0.711086とした。また、事前に入力する維持、処分、リスクなどのそれぞれにかかる費用や、図書館と保管庫の容量、本の大きさは正確な数値を図書館側から提供してもらえなかったため、仮の値を使用する。また、本の大きさや維持にかかる費用は本によって異なるが、今回の数理モデルに関しては全て同じ値を使用している。

- 図書館に本を置くための費用：100
- 保管庫に本を置くための費用：70
- 処分費用：30
- 本 1 冊の大きさ：3
- 図書館の容量：86076
- 保管庫の容量：86076
- 保管庫に本を置いたときのリスク費用：50
- 本を処分したときのリスク費用：200

図書館に本を置く場合の費用は、維持費のほかに人件費も必要となり、保管庫に本を置く場合は、運搬費と維持費が必要となる。これらの費用を比較すると、運搬費よりも人件費にかかる費用の方が大きいと考えられる。したがって、保管庫に本を置く費用よりも図書館に本を置く費用を大きい値としている。また、図書館と保管庫のそれぞれの容量については、まだ大学図書館と外部倉庫を合わせた収容スペースが限界でないことから、3 類の本すべてがどちらかには入りきる値から余裕を持たすために 1.2 倍し、その値を 2 等分している。また、使用するデータは第 3 類の各本が最後の 4 年間で借りられた冊数を用いている。

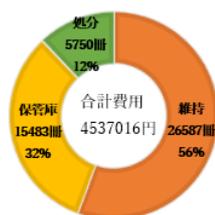


図 5 プログラムの実行結果

以上のデータでプログラムを実行したところ図 6 の結果となった。図 6 より、目的関数の値は 4537016 円となり、図書館に維持する本は 26587 冊、保管庫に移動する本は 15483 冊、処分する本は 5750 冊となった。これらの値は下限制約を余裕をもって超えるものである。

機械学習を用いた予測

第 3 類の本 1480 冊の予測の上限値を用いてプログラムを実行した。入力する定数は回帰分析のものと同じである。

図書館に 888 冊、保管庫に 591 冊、処分に 1 冊と、極端ながらも下限制約を遥かに超える結果となった。この場合は予測の上限値を取っている性質上、多くの本が 3 回以上借りられると予測されており、処分のリスク費用が大きくなりすぎてしまうため、処分には 1 冊しか割り当てられなかったのだろう。

回帰分析に基づいた結果では図書館に維持、保管庫に移動、処分にバランスよく割り当てられた。それに対して、

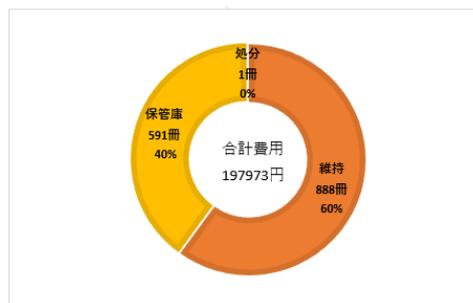


図 6 機械学習による計算結果

機械学習による結果は大半が図書館に維持と保管庫に移動に割り振られる極端なものになった。しかし、大学図書館側は本を処分するという行為を積極的に行えないようなので、そういった面から見れば、機械学習による計算結果の方が好まれるかもしれない。

また、これら 2 つの結果は一例であり、パラメータの数値を変更することで様々な結果を得ることもできる。

6 おわりに

本研究は、大学図書館に所蔵されている本の割り当てを、0-1 整数計画法の手法を用いることで解決するものである。単回帰分析と機械学習による結果を用い、2 つの割り当てシステムを制作した。それぞれで違った結果を得られるため、どちらのシステムが優れているかは実際に図書館側の意見を聞きたいと考えている。

また、まだシステムには改善点があり、機械学習による結果を出力するには千冊で約 1 時間を要するため、図書館全ての本の予測値を得るには途方もない時間を必要とする。予測値を出力する時間を削減することはできないため、今後図書館側が機械学習の結果を元に本の割り当てを行うことを考えるならば、何かしらの対策を講じなければならない。

さらに、単回帰分析や機械学習でデータを分析する際は、データを細分化した方がより詳しい結果が得られるだろう。

参考文献

- [1] Philip M. Morse: LIBRARY EFFECTIVENESS: A Systems Approach. The University of Chicago Press. 1969
- [2] 福嶋雅夫 新版 数理計画入門, 朝倉書店, 2013