

電子ジャーナル購読計画の数理最適化：九州大学における事例報告

畠塙, 晃平
九州大学附属図書館研究開発室 : 研究員

法常, 知子
九州大学附属図書館eリソースサービス室

<https://doi.org/10.15017/1669724>

出版情報：九州大学附属図書館研究開発室年報. 2015/2016, pp. 1-5, 2016-08. 九州大学附属図書館
バージョン：
権利関係：



電子ジャーナル購読計画の数理最適化 —九州大学における事例報告—

畠塙 晃平[†] 法常 知子[‡]

<抄録>

本報告では、電子ジャーナルの包括契約を雑誌毎の個別契約に切り替える際に生じる雑誌の選択問題を扱う。本事例報告では、梅谷[4]のアプローチに基づき、九州大学における 2014 年度の電子ジャーナルの購読・ダウンロードデータを用いて、電子ジャーナル購読計画の数理最適化を試みる。結果として、包括契約においてダウンロードされた論文を 7 割以上充足しつつ、かつ分野毎の公平性をある程度保証する個別契約の雑誌の組合せを求めることができた。

<キーワード> 電子ジャーナル、最適化

Mathematical Programming Approach for Electric Journal Subscription Planning —A Case Study at Kyushu University—

HATANO Kohei NORITSUNE Tomoko

1. はじめに

電子ジャーナルは学術情報基盤において不可欠な要素である。特に、人文系など一部を除いて、ほとんどの科学技術分野では、雑誌購読の形式が紙媒体から電子ファイルが主流となった。主要な契約の形としては従来は、個別の雑誌毎に購読契約を行っていたが、包括契約（いわゆるビッグ・ディール）が主流となっている。例えば、エルゼビア社の ScienceDirect などが挙げられる。包括契約により、プラスアルファの購読料金を上乗せすることで、個別に購読している雑誌群に加えてパッケージ内の個別購読していない雑誌群も購読可能になる。一方で、一旦包括契約を解除してしまうと、従来より個別購読している雑誌群のみがアクセス可能かつ個別購読していない雑誌群にはアクセスできなくなってしまう。

近年、ジャーナル価格は上昇し続けている。図 1 に分野毎の雑誌の平均価格をまとめた。2012 年から 2015 年の 3 年だけでも全分野平均で年 5% もの上昇率である。これは、3 年間で約 16% の価格上昇を意味する。

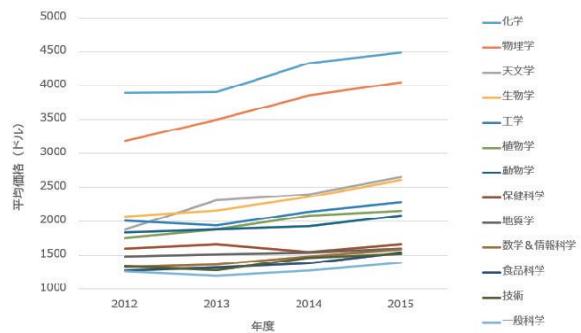


図 1:雑誌の平均価格（ Library Journal Periodicals Price Survey より）

しかもこの価格は米ドルで計算されており、円安の影響を考慮するとそれ以上の上昇率となる。ジャーナルの価格高騰の問題は尾城[5]や高橋[2]などで議論されている。

[†] はたの こうへい 九州大学附属図書館研究開発室研究員 (〒812-8581 福岡市東区箱崎 6-10-1) E-mail:
hatano.kohei.236@m.kyushu-u.ac.jp

[‡] のりつね ともこ 九州大学附属図書館 e リソースサービス室 (〒812-8581 福岡市東区箱崎 6-10-1) E-mail:
noritsune.tomoko.777@m.kyushu-u.ac.jp

電子ジャーナルの価格上昇への対処方法としては、機関リポジトリや OAジャーナルによるオープンアクセスの推進が考えられる。しかし、オープンアクセスの推進にはまだ多くの課題が残されており、今すぐ出来るというものではない。より短期的・直近の対策として、包括契約から個別契約への移行が挙げられる。実際、ジャーナル価格の高騰、円安の影響、さらに大学の予算削減により、包括契約から個別契約とペイパー方式(PPV、雑誌の論文1ダウンロード毎に料金を支払う方式)の組合せに移行する大学が増加している。2008年には、東邦大学が個別契約+PPVに切り替えることで、タイトル数をほぼ維持しつつ23%のコストダウンが達成されたとの報告があった[1]¹。他に個別契約に切り替えた大学をいくつか挙げると、2014年に名古屋大学(個別、ただし2015年に包括契約に復帰²)、中央大学、明治大学[3]、東京慈恵医科大学(個別+PPV)、滋賀医科大学(PPV)、2015年には大阪大学(個別)、岡山大学(個別)が個別契約または個別契約とPPVの組合せに移行した。2016年には佐賀大学が包括契約を中止している³。ただし、ここでのPPVは大学負担のPPVを指し、個人負担ではない。

包括契約から個別契約への移行の際、問題となるのは限られた予算内での購読雑誌の選択である。特に、複数の多様な学部の集まりからなる総合大学では、学部それぞれの思惑や事情があり、全体の調整は容易ではない。また、購読対象となる雑誌の候補は1000から2000以上におよび、人手による雑誌選択の作業はほぼ不可能といえる。

計算機に基づく個別契約の購読計画手法として、数理最適化によるアプローチが考えられる。梅谷[4]は個別契約の選択問題を数理計画問題の典型的な問題であるナップザック問題(およびその変形版)とし定式化した。その上で最適化問題ソルバを用いて、大阪大学の雑誌購読データ上で購読計画案を作成した。梅谷[4]の研究に代表される数理最適化のアプローチの利点は(1)透明性: 何を最大化すべきか目的関数が明確である、(2)効率性: 最適化ソルバの利用により、人手による作業が軽減される、(3)検証可能性: 誰でもデータとプログラムがあれば検証できる、の3点にある。特に、(1)の透明性は多様なステークホルダからなる総合大学においては望ましい性質である。

本事例報告では、梅谷[4]のアプローチに基づき、九州大学における2014年度の電子ジャーナルの購読・ダウンロードデータを用いて、電子ジャーナル購読計画の数理最適化を試みる。結果として、2014年度の雑

誌のダウンロード数を7.8割充足し、かつ分野毎の公平性をある程度保証する個別契約の雑誌の組合せを求めることができた。

表1:データの内訳

出版社	A社	B社	C社
雑誌数	1,711	1,677	1,389
ダウンロード総数	881,815	144,397	249,696

2. データ

九大における2014年度電子ジャーナル購読データ(A社、B社、C社)を用いた。表1にデータの概要をまとめた。各雑誌毎の基礎データは以下からなる:

1. ダウンロード数(DL数)
2. 分野種別
3. 個別契約した場合の価格
4. Impact Factor
5. 九大がその雑誌の論文を引用した回数

本研究では、ダウンロード数を需要の大きさとみなし、インパクトファクターや引用回数は考慮しない事とした。ただし、価格もしくはダウンロード数が不明な雑誌は除いた(約160誌)。また、過去に購読して購読中止後もアクセス可能な雑誌群も除いている。

3. 定式化と結果

本報告では梅谷[4]のアプローチに基づく数理最適化の枠組みを用いる。なお、最適化(後述)のプログラムはPythonで記述し、最適化ソルバはGurobi Optimizer 5.6を用いた。

本来、電子ジャーナルの購読計画は予測問題である。つまり、各年度初めに雑誌毎のダウンロード数をあらかじめ予測した上で購読計画を作成し、次に年度末にフィードバックとして実際のダウンロード数が与えられる、というやり取りで問題が定式化できる。しかし、本報告では、問題を簡単化するため、あらかじめ各雑誌のダウンロード数が与えられた下で、予算制約を満たしつつダウンロード数等を最適化するような購読計画を求める問題を考える。言い換えれば、予測問題ではなく、最適化問題として定式化する。

¹ <http://current.ndl.go.jp/node/30423>

² <http://current.ndl.go.jp/node/31265>

³ <http://current.ndl.go.jp/node/30782>

3.1. ナップザック問題としての定式化

まず、雑誌の選択問題をナップザック問題として定式化した。ナップザック問題とは数理最適化における典型的な問題の1つで、サイズの総計が与えられたナップザックを超えない範囲で、なるべく価値の総和の大きいアイテムの組合せを選ぶ問題である。サイズを雑誌の個別購読額、雑誌の価値をダウンロード数とみなすことで、雑誌の選択問題として以下のように定式化できる。

定式化1（ナップザック問題）

$$\max_{x \in \{0,1\}^n} \sum_{i=1}^n d_i x_i$$

subject to:

$$\sum_{i=1}^n c_i x_i \leq budget$$

ここで、 n は雑誌の個数、 $d_i (i=1, \dots, n)$ は雑誌 i のダウンロード数、 $c_i (i=1, \dots, n)$ は雑誌 i の価格、 $budget$ は予算上限、変数 $x_i \in \{0,1\}$ ($i = 1, \dots, n$)は雑誌を購読する/しないという選択を表す。簡単に言えば、この定式化は予算制約を満たしながら、雑誌を選択することで充足されるダウンロード数の総和を最大化する雑誌の組合せを選ぶ問題である。例えば、A社の場合、可能な雑誌の組合せは $2^{1711} \approx 10^{515}$ と膨大である。一般に、ナップザック問題は NP 困難という問題のクラスに属し、理論的な意味で効率的に解く手法は知られていない。しかし、高速な発見的手法や最適化ソルバが知られており、多くの応用例で高速に解くことができる。

定式化1を解いた結果を表2に示す。この結果はA,B,C社の雑誌群それぞれに対して、予算上限 $budget_A, budget_B, budget_C$ を与え、定式化1を解いた結果得られたものである。最適化で得られた雑誌の組合せにより、A,B社に対しては7割から8割程度ダウンロード数を充足していることがわかる。つまり、個別契約の組合せでも包括契約でのダウンロード数の7,8割は充足できるということである。しかも、選択された雑誌の割合は3割程度であり、少数の雑誌の選択でダウンロード数の多くが充足できることが確認できる。一方、C社に対しては、50%程度しかダウンロード数を充足できていない。これは、個別契約に切り換える場合、キャンパス毎に料金を支払わなければならず、九州大学は6キャンパスからなるので、価格が6倍になってしまふからと考えられる。

さらに、定式化1には次のような問題がある。図2にはA社の雑誌群に対する、分野毎のダウンロード数

表2: 定式化1に基づく結果

出版社	A社	B社	C社
充足された DL 数の割合 (%)	82	71	50
選ばれた雑誌の割合 (%)	35	24	4

の充足率を示す。図2を見ると、分野毎に大きな偏りがあることがわかる。例えば、化学分野のダウンロード数が 179,420 であるのに対して、数学分野のダウンロード数は 7,522 である。特に、数学や情報系の分野の充足率が 10%前後と低い。これは、定式化1においてはダウンロード単価（雑誌の価格/ダウンロード数）の高い雑誌が選ばれやすいためと考えられる。B社、C社についても同様の傾向がみられた（結果は省略）。

最適化によって充足されたダウンロード数の割合 (%)

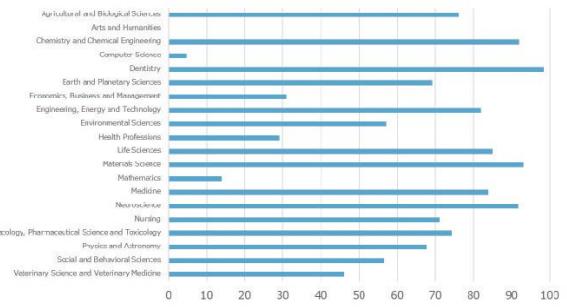


図2: 定式化1を解いた際の分野毎のダウンロード数充足率（A社）

3.2 制約付きナップザック問題に基づく定式化

定式化1の問題点は分野毎のダウンロード充足率を考慮していない点にある。そこで梅谷[4]は分野毎のダウンロード充足率の下限値を最大化する次のような定式化を提案した：

定式化2（分野均等化制約付きナップザック問題）

$$\max_{x \in \{0,1\}^n} z$$

$$x \in \{0,1\}^n$$

subject to:

$$\begin{aligned} \sum_{i \in N_j} d_i x_i &\geq_z D_j \quad (j = 1, \dots, k) \\ \sum_{i \in N} c_i x_i &\leq budget \end{aligned}$$

ここで、 N を雑誌の集合、 $n=N$ 、分野数を k とし、 $j = 1, \dots, k$ に対して N_j を分野 j における雑誌の集合、 D_j を分野 j に属する雑誌のダウンロード総数と

表 3: 定式化 2 に基づく結果

出版社	A 社	B 社	C 社
充足された DL 数の割合 (%)	78	67	39
選ばれた雑誌の割合 (%)	39	28	6
分野別最低充足率 (%)	78	67	32

する。つまり、定式化 2 では予算上限を超えない範囲で分野毎の充足率の下限を最大化する雑誌の組合せを求めるのである。ただし、この定式化を解くだけでは予算上限を大きく下回る組合せが output されるおそれがある。そこで、得られた最低充足率 z を満たしながら、選択された組合せのダウンロード総数を最大化する組合せを改めて求める（2段階の最適化）。

定式化 2 による最適化の結果を表 3 にまとめた。定式化 1 と比較すると、A,B 社に対しては、ダウンロード総数を数 % 低下させつつも分野毎の最低ダウンロード充足率を約 7、8 割に高めており、分野毎の偏りは解消されている。しかし、C 社のデータに対しては、やはり、価格が 6 倍になっているという条件は大きく、ダウンロード総数を約 10% 低下、分野別最低充足率は 3 割程度にとどまっている。図 3 に A 社のデータに対して定式化 2 を解いた際の分野毎のダウンロード充足率を示す（B 社、C 社も同様の結果が得られた、省略）。

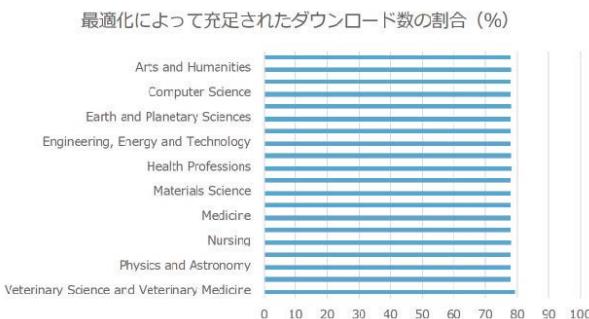


図 3: 定式化 2 を解いた際の分野毎のダウンロード数充足率 (A 社)

3.3. 個別契約の組合せ + PPV

最後に、梅谷[4]では考慮されていなかった PPV を取り入れた定式化を考える。本報告では単純化のため、各雑誌において、個別購読する/しない/すべて PPV で購読の 3 択のみを仮定する（つまり、部分的に PPV で賄うといったケースは考えない）。この設定の下では、 $(\text{購読価格}/\text{ダウンロード数}) > \text{PPV 価格}$ を満たす雑誌は、PPV で購読した方が得になる。よって、購読価格を PPV 価格 * ダウンロード数で置き換えることにより、定式化 2 と同じ枠組みで扱える。

表 4: 定式化 2 + PPV に基づく結果

出版社	A 社	B 社	C 社
充足された DL 数の割合 (%)	78	67	42
選ばれた雑誌の割合 (%)	42	44	43
分野別最低充足率 (%)	78	67	37

結果を表 4 に示す。ダウンロード数の充足率については A,B 社については変化はなく、C 社に対して 39% から 42% に増加した。選択された雑誌の割合は A,B,C 社それぞれに対して、39% から 42%，28% から 44%，6% から 43% に上昇した。また、分野別最低充足率は C についてのみ 32% から 37% に上昇した。PPV を考慮することにより、少数のダウンロード数の充足に貢献しているが、残念ながらダウンロード数の充足率について大きな向上はみられない。これは、最適化により、ダウンロード単価の小さい雑誌群が予め優先して選ばれるために、PPV への切換えによる貢献の余地が小さいためと考えられる。

4. まとめと今後の課題

本報告では、梅谷[4]の数理最適化のアプローチに基づき、九州大学の電子ジャーナルのダウンロードデータを用いて電子ジャーナルの雑誌選択問題の最適化を試みた。結果として、2014 年度の雑誌のダウンロード数を 7.8 割充足し、かつ分野毎の公平性をある程度保証する個別契約の雑誌の組合せを求めることができた。また、本結果は、梅谷の結果[4]と同様を傾向を示している。

しかし、多くの問題が残されている。1 つは問題の定式化そのものの妥当性である。本報告ではダウンロード数を需要そのものとみなしたが、インパクトファクターなど他の尺度を考慮する事も可能である。また、本報告では分野間の公平性を担保する制約条件を考えたが、総合大学においては部局間の公平性を考慮に入れなければならない。また、本来、電子ジャーナルの選択問題は予測問題として定式化するのが自然である。機械学習手法を援用した購読計画は興味深い課題である。

謝辞

本報告の作成にあたり、大阪大学の梅谷俊治氏には様々な情報を提供していただいた。この場を借りて感謝申し上げたい。

参考文献

- [1] 吉田杏子 .東邦大学における外国雑誌価格高騰への対応 .*薬学図書館* ,57:25-30, 2012.
- [2] 高橋努 .大学図書館から見た電子ジャーナルの現状と課題 .*電子情報通信学会誌* ,95(1):27-32, 2012.
- [3] 仲山加奈子 and 菊池亮一 .*ビッグディールのおわりとこれから .図書の譜 :明治大学図書館紀要* ,18:227-239, 2014.
- [4] 梅谷俊治 .電子ジャーナル購読計画の効率的な作成. In 日本オペレーションズ・リサーチ学会 2015 年秋季研究発表会 ,2015.
- [5] 尾城孝一 .*ビッグディールは大学にとって最適な契約モデルか? SPARC Japan News Letter*, 5, 2010.