

データサイエンス教育において 「数理モデリング」をいかに教えるか

木庭 淳^{1,a)} 加藤 直樹^{1,b)}

概要: 本稿は、データサイエンス系の学部である兵庫県立大学社会情報科学部において、必修科目「数理モデリング」を実際に講義するにあたり、その目的、カリキュラムの中での位置づけ、テキスト選定の過程を述べた上で、どのようなテーマ内容で授業を行ったかを報告する。問題解決の手段としてどのように問題を定式化するかという部分は説明が難しく、ともすれば一般化の観点から抽象的に扱われたり、あるいは理論的にのみ議論されたりすることが多い。従って初心者向けの授業でどのように講義すべきかという問題は正解に到達することは難しい。今回は初めての試みとして、検討すべき点も含めて報告する。

How to Teach “Mathematical Modeling” in Education of Data Science

1. はじめに

データサイエンス教育において数理モデル化技術の修得は大きな位置を占めている。すなわち現実の問題の本質をとらえ、数理的な解決の糸口をつかむためには、どんな仮定のもとで何を変数として最適化を行うのか等を明確にする必要がある。これらのノウハウを得るためには、高校までの授業ではそれほど意識しなかった考え方を含めて、修練を重ねていく必要がある。これが現実の問題に出会ったときに論理的に考えを組み立て、他人に納得してもらえようような妥当性のある解を導き出す基礎になると考えられる。

高校までの数学においては、入学試験や定期試験を意識した問題作りとなっているため、唯一の解答にたどり着くための問題設定や仮定が明確になっていなければならない。今まで数学といえばこのような問題文が与えられ、条件が与えられているものと思い込んでいた学生が、そのような問題文もなく、条件も与えられていない状況に出会ったときに急に不安になるようだ。モデリングの授業においては、問題を解きやすいように、あるいは現実の解に近くなるように、自分で変数を決め、仮定を決めていくことが求められる。これが実際の問題を解決するために必要な訓

練であると考えている。それを分らせるために、授業としてはどのような資料と解説を与え、どのような練習をさせればよいか、というのが我々の主要なテーマである。

本報告では、数理モデリングの技術をいかに教えるか、というテーマでの教育の試みを報告する。学生層は一律ではなく、前提となる知識を彼らが修得しているとは必ずしも期待できない。この環境のもとで、問題解決のための予備知識をある程度解説しつつ、それをベースにモデリングの実例を示していく必要がある。さらに一定時間後にもう一度聞いたとき、「その話、どこかで聞いた覚えがある」とかすかに記憶に残っているレベルではなく、自分で実際にモデリングをすることができるようにさせたい。そのようにある意味欲張った目標のもとで本授業を実施した。

この目標を達成するために、我々が今回の授業で力を入れたことは、典型的なモデリングについて具体的なイメージを持ってもらう、ということである。そのために、できるだけ自分でモデル化の作業をしてもらうように誘導した。すなわちほぼ以下の手順に沿って授業を行った。

- (1) 具体的かつ簡単な数値例を例題とした。
- (2) 例題に対する予備知識・解法の解説を述べた後、数値を変えただけの小問を解いてもらった。
- (3) 時間的に余裕があれば小問より応用的な練習問題を解いてもらった。
- (4) 余裕がなければ課題とし、次週の時間初めに解説した。

¹ 兵庫県立大学社会情報科学部
GakuenNishi, Nishi-ku, Kobe 651-2197, Japan

a) kiniwa@sis.u-hyogo.ac.jp

b) naoki.katoh@sis.u-hyogo.ac.jp

本章以降の構成は以下の通りである。第2章では我々が置かれている環境と「数理モデリング」という科目の位置づけを述べている。第3章では教材の内容はどの分野なのか、それらはどのようにリンクされているのか、いかにテーマの取捨選択をしたのか、に焦点を当てている。第4章は教え方や授業資料の特徴等についての記述である。最後に第5章で結論を述べる。

2. 背景

ここでは昨今のデータサイエンス系学部新設に関して兵庫県立大学における特色、在籍する学生層の状況、「数理モデリング」という科目がカリキュラムの中でどのような位置にあるかといった背景について述べる。

2.1 データサイエンス教育

滋賀大学のデータサイエンス学部を皮切りに全国の大学でデータサイエンス系の学部が開設され、データサイエンティスト養成の試みが実施されつつある。国内ではデータを統計的に分析する流れが多いと思われるが、こうした中で兵庫県立大学におけるデータサイエンス教育の特徴は、

- 主に情報科学的な手法でビッグデータの分析を行うこと
- 企業と提携した現場に即した授業を展開して学生のモチベーションを高めること
- PBL 演習など学生が積極的に問題解決に携わる実習を通して知識を生かせる方法を身につけさせること

にあると考えている。兵庫県立大学においては、前身のひとつである神戸商科大学管理科学科の流れを汲んでおり、その卒業生たちが現在も関西の企業・大学に勤務しているため、伝統や人材等の資源を生かした教育が行われている。

2.2 社会情報学部と在籍する学生層

兵庫県立大学社会情報学部は2019年4月より新入生の受け入れを開始した。1学年の定員は100名で、現在1、2年次の学生200余名が在籍している。そして彼らのうち6~7割が理系である。高校時代に数IIIを学習していない学生のために、1年次において「社会情報科学のための数学」という必修科目を設けて、数学的基礎を身につけてもらうようにしている。しかしながら文系の学生と理系の学生では、数学的思考そのものに対する考え方かなりの差があるようで、どうしても数学への苦手意識から抜け出せない学生が見受けられることも事実である。一方でビッグデータやAIなどへの社会的な関心の高まりを受けて、学部のカリキュラムを積極的にこなすのみならず、データサイエンスに関するシンポジウムに参加したり、勉強会をつくってグループで議論する学生たちも存在する。大学院への進学に興味を持つ学生も、3分の1程度はいるようである。

2.3 必修科目「数理モデリング」の位置づけ

「数理モデリング」という科目は、社会情報科学部の学生が2年次前期で修得すべき必修の専門教育科目として位置づけられている。従来より大学教養科目の最初に教えられる微積分・線形代数・確率統計の次に配置されている(図1)。つまり、社会情報科学部の専門性を特徴づける基本的な科目という認識であろう。学生は「数理モデリング」を履修後、2年次後期の「オペレーションズ・リサーチ」で線形計画法について、同じく2年次後期の「多変量解析」で重回帰分析について理論的側面を学ぶ。さらに3年次の「グラフ理論」や1年次の「確率・統計」とも若干のつながりがある。これより履修後の延長線が応用的な専門科目に結び付くような基礎固めの位置にある、と言える。言い換えると本科目の果たすべき役割は、理論的な追求よりも現実問題にどのように応用できるのかという、基礎から応用への橋渡しを担うべきではないかと考えられる。

3. 「数理モデリング」のテーマ内容

あらかじめ講義内容については大筋の柱は設定していたが、具体的な内容を決めていくにあたり、調査を含めて約半年前から入念に内容の検討を始めた。ここでは、授業で扱うテーマの構成と各章の関係、理論的説明やツールの取捨選択の判断について述べる。

3.1 全体の構成と各章の関係

最初の計画段階から考えていた大筋の構成は図2の通りである。矢印は関連する話題(根元の方が基礎部分)である。線形計画法から始まる一連のテーマは、線形計画法の関連分野がいかに広いかを示すという意味がある。最後の「統計的モデリング」と「シミュレーション」は他の章との関連は薄い、データサイエンスには統計的な分析が欠かせないこと、事象が複雑すぎる場合にシミュレーションの知識は欠かせないことより、この二つをとりあげた。

また第1回目の授業時間には、「数理モデリング概説」として一般論を述べた。これはもし最初から詳細な議論をスタートしたとすると、おそらくかなりの学生はこの授業がいったい何の役に立つのか、自分が今習っている理論に何の意味があるのかが分からなくなるのではないかと、ということ予想したためである。そこでまず初回において、数理モデリングがいかに実社会に役に立っているかを

- (1) 最近多発している災害に際しての避難計画、および
 - (2) 新型コロナウイルスに対する数理モデリング
- という実例をあげて概説した。

目的地あるいは到達目標を示したうえで、各論の詳細を解説していくという方法をとった。

3.2 テーマおよびツールの取捨選択

授業に含めるべきかどうかを検討するに際して、主に問

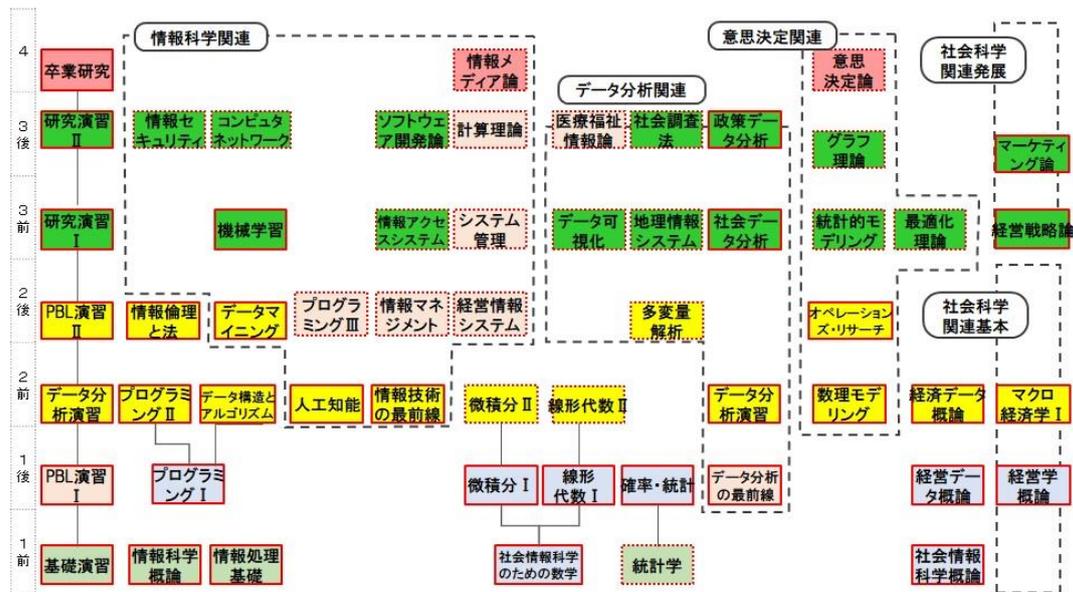


図 1 履修モデル [6]

Fig. 1 A course option

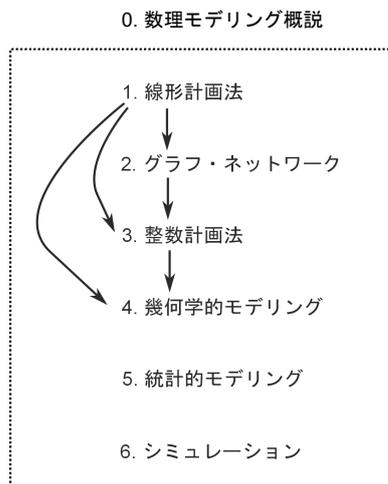


図 2 全体の構成

Fig. 2 Configuration

題となったのは次の二点である。

- (1) どの程度まで理論的な説明をすればよいか。
- (2) モデル化だけでよいか、定式化が終わった後に解を求めるにはどうするか。

(1) の理論的な説明について特に検討を要したのは、以下のいくつかのテーマにおいてである。

単体法

線形計画法における単体法の説明について、予想される難点としては

- 基底解、ピボット演算、スラック変数など専門用語が多く、全員に概念を理解させるのは難しいのではないかと。また説明のために時間がとられてしまい、本来のテーマであるモデリングの部分に充分時間がとれないのではないかと。

- 一般的な数式を使って説明すると、数式の意味を把握することが難しいのではないかと。

といった点である。しかし、定式化できた後に解をソルバーで求めたとして、なぜ解が求まるのかがまったくブラックボックスであるのも問題であるとも考えられる。そこで今回とった方法としては、

- 単体法の理論的な説明は避ける。
- 単体法を直観的に理解してもらうために、高校で学習した平面領域での線形関数の最大・最小を求める方法を例示する。
- 3次元以上の制約式の場合は平面領域の方法から類推してもらう。

整数計画法の解法

次に理論的な説明をどうすべきか検討したのは、整数計画法における切除平面法と分枝限定法についてである。これらも一般的に説明すると、おそらくほとんどの学生が理解不能となり、その後の授業が続けられなくなる恐れがある。

逆に、なぜ整数計画法を解くことができるのかにまったく触れないということは、学生の素朴な疑問や興味に答えることができず、不透明な理解のまま授業を終えてしまうことになりかねない。

そこで我々がとったのは、単体法の場合と同じく視覚的に直観に訴える方法である。切除平面法については図3によって理解してもらうことにした。また分枝限定法については資料としては準備せず、口頭で簡単に説明することにした。

回帰分析の概念

最後に苦心したのは統計的な基礎知識を前提とする回帰

切除平面法

切除平面法についてアイデアのみを簡単に解説する。以下の例を考える。制約を満たすのは白丸で示した格子点となる。このとき x_i が整数である条件を無視して LP を解いても、最適解は得られない (図 3(a))。そこで不等式の範囲内にある格子点を囲むように点線で示した範囲を考える。ちょうど格子点の場所に釘(くぎ)があると考え、その釘を囲むように輪ゴムをひっかけると思えばよい。これらの点線を制約式で表した上で単体法を用いて解けば、最適な整数解が得られる (図 3(b))。

$$\begin{aligned} & \text{maximize } x + y \\ & \text{subject to } 6x_1 - 5x_2 \geq 11 \\ & \quad 3x_1 + 10x_2 \leq 54 \\ & \quad x_i \text{ は整数} \end{aligned}$$

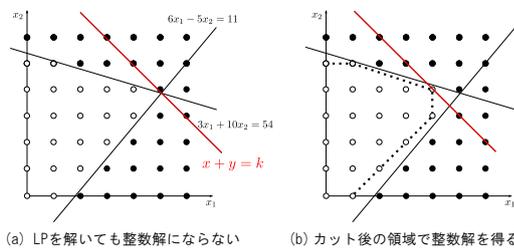


図 3 切除平面法

Fig. 3 A cutting plane method

分析の概念についてである。例えば

- 相関係数が -1 以上 1 以下であることを、原点が平均となる二種類のベクトル x, y の内積すなわち $\cos \theta$ と等価であることから説明。
- 最小二乗法を具体例で偏微分を使って説明。
- 決定係数において残差の変動と実測値の変動の違いを図で説明。

など、回帰分析や重回帰分析を行うためには予備知識として知っておかなければならない事項を、どう理解させるのが問題であった。これらの説明にあまりに時間を費やすと、モデリングの説明が十分にできない恐れがある。

そこでこれらはまとめて付録として資料に収め、最初に軽く説明することで後から随時復習をしてもらうことにした。1年次の後期で確率統計の授業があり、一般的な説明は受けているということが安心材料となった。

次に(2)の問題については、最適解等が実際どんな値になるのかは気になるところだと思われる。定式化まで終了して、その後は「もし高価なツールを購入することができれば解けるはずだが、お金が無いので無理」であるとか、「このツールを使えば解けるはずだが、使い方が複雑すぎてその説明に1時間かかる」というのでは、何のために定式化したのか分からず、欲求不満がたまってしまう。

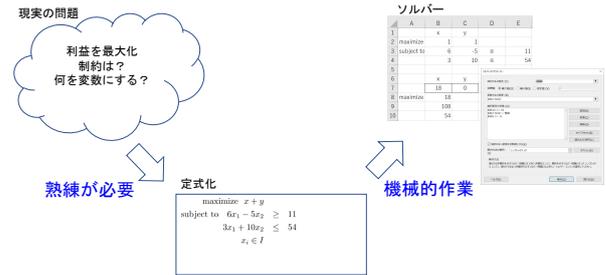


図 4 問題→モデル化→求解

Fig. 4 Problem → Modeling → Solving

そこで簡単な例題について可能な限り必ず解を求める*1、という方針のもとに解を得るためのツールを探した。適切なツールの条件について、

- 誰でも手軽に入手できるものがあれば、それを利用する。
- できるだけひとつのツールで多くの問題が解けるものがよい。

という観点から探した結果、線形計画法と整数計画法には Excel ソルバー、回帰分析には Excel のデータ分析ツール、シミュレーションには Excel のワークシートを使うことにした。いずれも Exce にアドインを組み込む等すれば使用可能になる。また幾何学的モデリングでポロノイ図を扱ったが、これは「ひなた GIS[4]」に組み込まれている機能を利用させていただいた。

4. 授業方法と学生の理解

本章では授業での教え方の工夫をどのようにすべきかを、文献等の例をまじえて述べる。また授業後の学生の反応をまとめる。

4.1 「モデリング」をどう教えるか

問題解決のときにたどる道筋は一般的にはおそらく図 4 のような図式で理解され、現実に関何かがあったとき、それを定式化またはモデル化した上で、プログラムによってコンピュータで解決する、という流れになる。かつてはプログラム化することに多大な努力・熟練が必要であったと思われる。そこでこれを何とか簡単に自動化できないかということで、ソルバーというソフトが開発されるようになり、定式化さえできればある程度機械的作業で求解まで進められるようになった。しかしながら、現実の問題からモデル化に至る過程の自動化が実現することは今後もなさそうである。従って、データサイエンティストとして育つ人材も、この部分には習熟しておく必要があると考えられる。

*1 巡回セールスマン問題については、例外的に定式化の説明のみ。

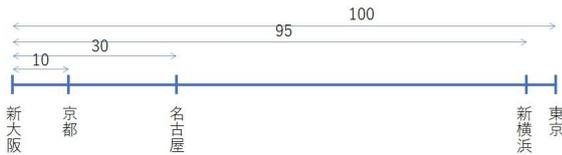


図 5 駅間距離

Fig. 5 Distances between stations

数理モデリングをどのように教えるかという試みは、今までにも何度もなされてきたと思われる。例えば文献 [7] においてはこのつかみよしの無い過程を分析して、5つのステップで解決する方法論を展開している。すなわち図4の「問題」→「モデル化」の間に「数理モデルの手法を選ぶ」というステップを追加し、「求解」の後に「(モデルの解ではなく) 問題の解答を出す」というステップを追加する。我々の方法も基本的には大差はない。

また文献 [3], [8] では特に確率分野の数理モデリングを読み物風に読みやすく記述されている。学生には夏季休業の課題で自習用として使うには適しているかもしれない。いずれの文献もカリキュラム内での他科目との接続性、難易度等の理由から、今回のテキスト採用からは見送った。

文献 [1] は様々な数理モデルが図を使用して直観的に分かりやすく並べられている。事典的な役割として使うには役に立つが、本授業のような修練という目的からは少しはずれると判断した。

4.2 学生の理解をあげるための工夫

授業資料の構成

1年間の半期15週という限られた時間の中で、理論的な側面をほとんど知らない学生全員に対して、どうしても限られた分野の中から限られた方法で授業をせざるを得ない。現実の問題に対してどのようにモデル化するかという工夫については様々な種類のモデル化(例えば理論的なモデル化やシミュレーションによるモデル化等)が可能であろうと思われる。ただこの種の問題に対してはこのパターンの解法、というイメージのようなものができあがっていれば、実際の現場に臨むときには取り組みやすいであろう。

そこで、そのようなイメージが頭の中に形成されやすいように、月並みではあるが

- (1) 例題
- (2) 解法のための予備知識
- (3) 例題に対するモデルの作成
- (4) ソルバー等求解のためのツールの説明
- (5) 練習問題

の順に、問題ごとにパターン化した構成とした(付録参照)。

複数のモデリング例

次に様々なモデリングがあり得るということを理解してもらうために、いくつかのモデル化の変種を示した。

例題 [5]

ある会社の大阪支社の社員30人と東京支社の社員15人が一か所集まって会議をすることになった。交通費の総額を最も少なくするためには、どこに集まればいいか。

この例題に対し、大阪をP、東京をQとして、定式化1のために次の仮定をおく。

仮定

- (1) PQ間の距離を100kmとして考える
- (2) 運賃は移動距離に比例する
- (3) 単純に距離で答えを出し、ちょうど駅の位置でなくてもいいものとする。

変数 x kmを地点Pからの距離として式で表すと、全員の総移動距離は

$$30x + 15(100 - x) = 15x + 1500$$

となるので、定式化1は

定式化1

$$\begin{aligned} &\text{minimize } 15x + 1500 \\ &\text{subject to } x \leq 100 \\ &\quad \quad \quad x \geq 0 \end{aligned}$$

この定式化1の導出過程を示した上で、学生には以下のような課題を与えた。

課題

この会社の支店が新幹線の東京、新横浜、名古屋、京都、新大阪の駅の近くにあるものとする。このときこれらの駅のいずれかを集合場所とするということにして(上記仮定(3)を差し替える)、定式化2を行いなさい。ただし、各駅間の距離は図5とする。

つまり問題が与えられたとき、仮定を変えることによって別のモデルを得ることができる、ということを学生に認識してもらうための課題である。定式化2の解答例は以下のようなになる。

解答例(定式化2)

$$\begin{aligned} &\text{minimize } 15x + 1500 \\ &\text{subject to } x \in \{10, 30, 95, 100\} \end{aligned}$$

4.3 学生の反応

授業後の学生の反応は以下の通り。ほぼ半分の49名が回答している。理解が難しいという意見もあったが、概ね肯定的に評価するとの回答を得てほっとしている。

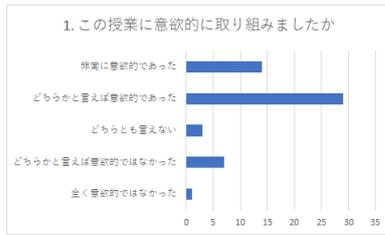


図 6 授業への意欲
Fig. 6 Incentive to the lecture

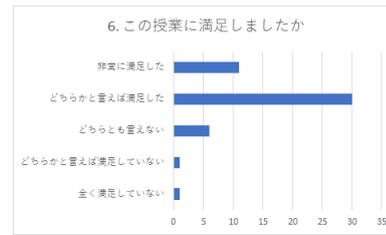


図 11 満足度
Fig. 11 Satisfaction level

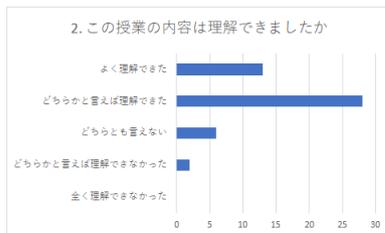


図 7 理解度
Fig. 7 Understanding level

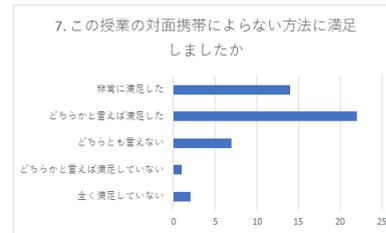


図 12 オンライン環境への満足度
Fig. 12 Evaluation of the online environment

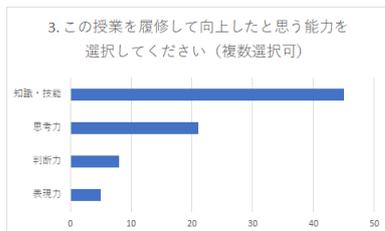


図 8 得られた能力
Fig. 8 Gained ability

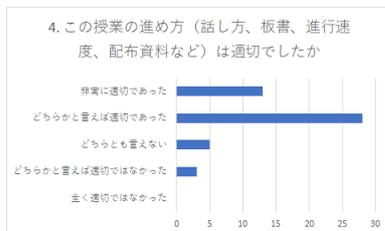


図 9 授業の評価
Fig. 9 Evaluation of the lecture

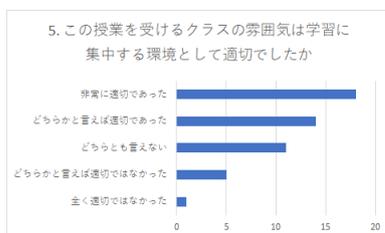


図 10 集中できる環境
Fig. 10 Concentration

第 1 の反省点としては、複数のモデリング (第 4.2 章) の可能性があることは第 1 章でとりあげたが、各章 1 つずつくらい例示すればよかったのではないかと考えている。ただし問題によっては難しくなったかもしれない。

第 2 に、学生自身に問題を作らせるという方法もあったかもしれない。何が問題なのかを考えて変数や条件を設定する部分は、高校まではまったく経験がないことなので、かなり混乱が予想されるが、非常に重要な部分である。

第 3 に、今回は新型コロナ禍によるオンライン授業という特殊な形態で、思うようにきめ細かなサポートができなかったのが残念だった。

謝辞 本研究は兵庫県立大学特別研究助成 (先導研究 A) を受けた。また授業資料に ascolorbox.sty を使用した。

参考文献

- [1] 江崎貴裕、「データ分析のための数理モデル入門」、ソシム (2020)
- [2] 加藤直樹、「数理計画法」、コロナ社 (2007)
- [3] 浜田宏、「その問題、数理モデルが解決します」、ベレ出版 (2018)
- [4] 地理情報システム「ひなた GIS」、宮崎県情報政策課 (2017), <https://hgis.pref.miyazaki.lg.jp/hinata/>
- [5] 平本巖、栗原和夫、「文科系の線形計画法入門」、牧野書店 (2000)
- [6] 兵庫県立大学社会情報科学部、「2020 年度履修の手引き」 (2020)
- [7] M.Meerschaert、佐藤一憲、梶原毅、佐々木徹、竹内康博、宮崎倫子、守田智、「数理モデリング入門—ファイブ・ステップ法—」、共立出版 (2015)
- [8] 水上ひろき、熊谷雄介、高野雅典、藤原晴雄、「データ活用のための数理モデリング入門」、技術評論社 (2020)

5. 今後の課題・検討

最後に今後の課題等について考察する。

付 録

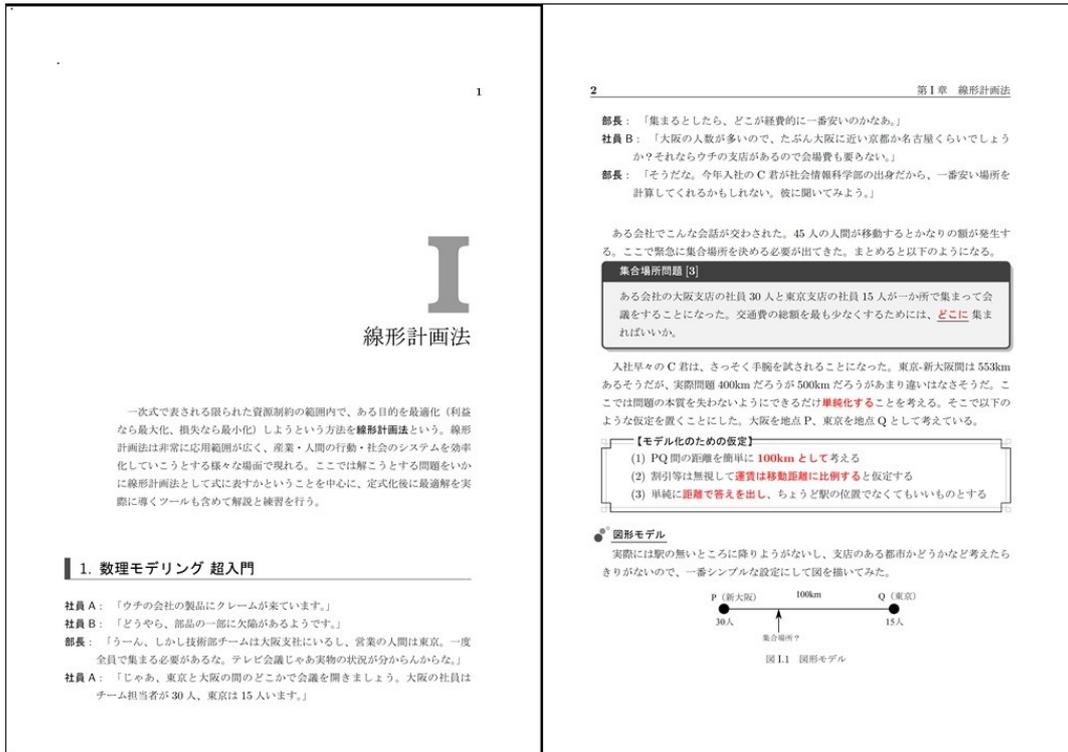


図 A-1 線形計画法 (pp.1-2)
Fig. A-1 Linear Programming (pp.1-2)

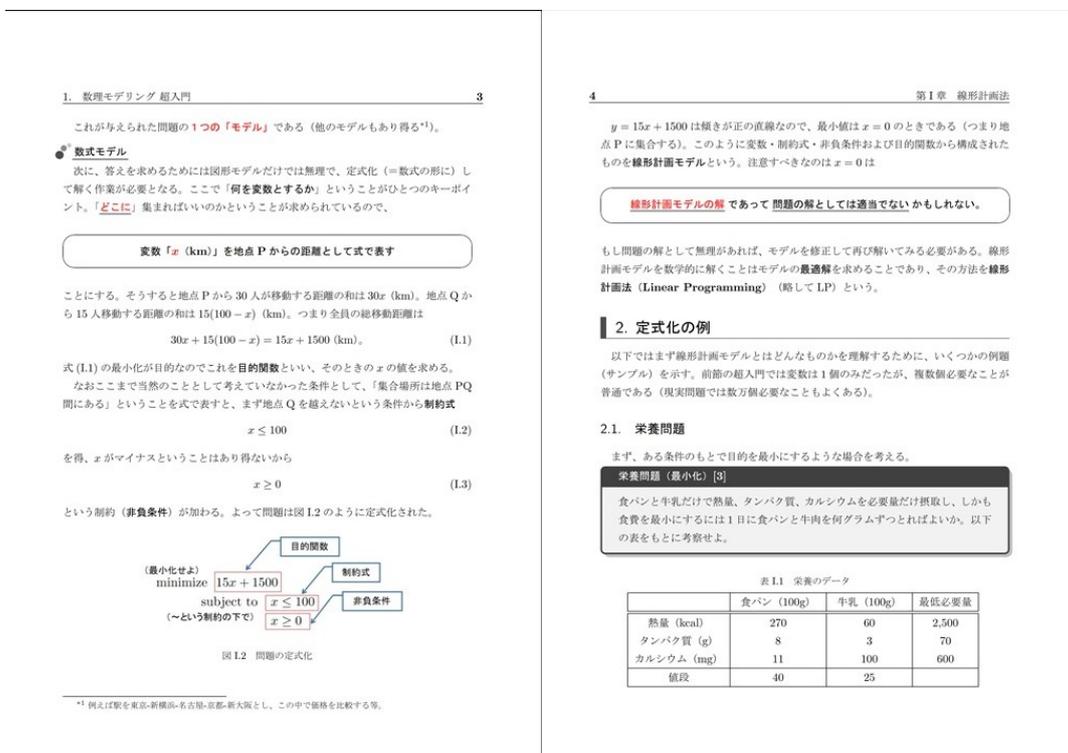


図 A-2 線形計画法 (pp.3-4)
Fig. A-2 Linear Programming (pp.3-4)